

CONTINUOUS AVERAGING PROOF OF THE NEKHOROSHEV THEOREM

JINXIN XUE
UNIVERSITY OF MARYLAND-COLLEGE PARK

ABSTRACT. In this paper we develop the *continuous averaging* method of Treschev to work on the simultaneous Diophantine approximation and apply the result to give a new proof of the Nekhoroshev theorem. We obtain a sharp normal form theorem and an explicit estimate of the stability constants appearing in the Nekhoroshev theorem.

1. INTRODUCTION

In the papers [Tr1, Tr2], Treschev developed a new averaging method called continuous averaging. It is a powerful tool to derive sharp constants in the exponentially small splitting problems in Hamiltonian systems with one and a half degrees of freedom. But the technicality becomes very heavy when we use the method to study Hamiltonian systems of more degrees of freedom. For this reason, the method has not been applied to other problems yet.

In this paper, we use the continuous averaging to give a new proof of the Nekhoroshev theorem. We consider the following analytic nearly integrable Hamiltonian system:

$$(1.1) \quad H(I, \theta, x, y) = H_0(I) + \varepsilon H_1(I, \theta, x, y),$$

The phase space is

$$(I, \theta, x, y) \in \mathcal{D} := \mathcal{G}^n \times (\mathbb{R}/2\pi\mathbb{Z})^n \times \mathcal{W}^{2m} \subset \mathbb{R}^n \times (\mathbb{R}/2\pi\mathbb{Z})^n \times \mathbb{R}^{2m}, \quad n \geq 2, \quad m \geq 0.$$

We complexify the variables and extend the domain of (I, x, y) to a σ neighborhood and that of θ to a ρ neighborhood of the original domains respectively. The extended phase space to the complex domain is

$$\mathcal{D}(\rho, \sigma) := (\mathcal{G}^n + \sigma) \times ((\mathbb{R}/2\pi\mathbb{Z})^n + \rho) \times (\mathcal{W}^{2m} + \sigma) \subset \mathbb{C}^n \times (\mathbb{C}/2\pi\mathbb{Z})^n \times \mathbb{C}^{2m},$$

where ρ is the width of analyticity in θ and σ is that of the slow variables I, x, y .

As stated in [Ne, L1, L2, LN, LNN, Po, BM], Nekhoroshev Theorem ensures that when the unperturbed Hamiltonian H_0 is quasi-convex, by which we mean that

Email: jxue@math.umd.edu .

the set $\{I \mid H_0(I) \leq E\}$ is strictly convex, the following general estimate holds for sufficiently small ε :

$$(1.2) \quad \|I(t) - I(0)\| \leq C_0 \varepsilon^b, \quad \text{when} \quad |t| \leq \mathcal{T} = C_1 e^{C_2/\varepsilon^a}.$$

for some constants $a, b, C_0, C_1, C_2 > 0$ independent of ε , where $I(t)$ is the action variable component of any orbit associated to Hamiltonian (1.1) with initial condition in the set \mathcal{D} .

There are many works studying the stability exponents a and b (c.f. [LN, Po, BM]). Their approaches are based on a careful study of the geometric and number theoretical aspects of resonances. Instead, in this paper we try to sharpen the estimates in the analytic part of the proof using continuous averaging to obtain an improved normal form (see Theorem 3.1). Then we apply the normal form to Lochak's argument to get the Nekhoroshev theorem (see Theorem 2.1) where all the stability constants are estimated explicitly. In this paper, we only work on the case $a = b = 1/2n$. But the normal form theorem can be easily applied to other prescribed a and b to get the corresponding C_2 .

The method of Lochak is called the simultaneous Diophantine approximation, which turns out to be an important alternative to the classical approach via small divisor techniques, as explained in [L2]. Its main idea is to do the averaging in a vicinity of a periodic orbit. So it is essentially an averaging procedure for systems with one fast angle. In general, we can kill the dependence on the fast angle up to exponential smallness. This makes the simultaneous Diophantine approximation suitable to prove the Nekhoroshev theorem. The work [PT] can be considered as a development of the continuous averaging to the small divisor case. In this paper, it is the first time that the continuous averaging has been developed to the simultaneous Diophantine approximation.

We point out the relation between continuous averaging and some important PDEs. The idea of the continuous averaging is to study the averaging procedure using PDE instead of iterations. The PDE has the form $H_\delta = \{H, F\}$, where F is the Hilbert transform of H in some special cases (see Section 3 for more details). This type of equation has been studied (c.f. [CCF]) as a simplified model for quasi-geostrophic equation (c.f. [KNV]), incompressible Euler equation, etc. It would be interesting if we could apply some PDE techniques to our problem.

To state our theorems, we need the following definitions.

Definition 1.1. (1) We use $|\cdot|$, $|\cdot|_2$, $|\cdot|_\infty$ to denote the l_1 , l_2 , l_∞ norms for a vector in \mathbb{R}^n or \mathbb{Z}^n . Without causing confusion, we also use $|\cdot|$ to denote the absolute value of a function whose range is in \mathbb{R} or \mathbb{C} .

(2) For a function $f(I, \theta, x, y)$, the weighted Fourier norm is defined as:

$$\|f\|_{\rho'} = \sup_{I, x, y} \sum_{k \in \mathbb{Z}^n} |f^k| e^{|k|\rho'}, \quad \rho' \leq \rho,$$

if we have the Fourier expansion $f(I, \theta, x, y) = \sum_{k \in \mathbb{Z}^n} f^k(I, x, y) e^{i\langle k, \theta \rangle}$, and the other variables $(I, x, y) \in (\mathcal{G}^n + \sigma) \times (\mathcal{W}^{2m} + \sigma)$.

(3) $\max\{\|H_1\|_\rho, \|\nabla H_1\|_\rho\} := \mu$.

We also use the following definition to characterize the convexity of the unperturbed part $H_0(I)$.

Definition 1.2. Consider a Hamiltonian $H_0(I)$ defined on $G^n + \sigma$. Then, we define the associated constants $M_\pm > 0$ to characterize the convexity of $H_0(I)$.

$$(1.3) \quad \begin{aligned} 0 < M_- |v|_2^2 &\leq |\langle \nabla^2 H_0(I) v, v \rangle|, \quad v \in \mathbb{R}^n \setminus \{0\}, \quad I \in \mathcal{G}^n. \\ |\nabla^2 H_0(I) v|_2 &\leq M_+ |v|_2, \quad v \in \mathbb{C}^n \setminus \{0\}, \quad I \in \mathcal{G}^n + \sigma. \end{aligned}$$

Now we state a simplified version of our main theorem. The complete version is stated in the next section.

Theorem 1.1. Consider a Hamiltonian system (1.1) satisfying inequalities (1.3) in Definition 1.2 and $n \geq 2, m \geq 0$. For every orbit $(I, \theta, x, y)(t)$ with initial condition $(I, \theta, x, y)(0) \in \mathcal{D}(\rho, \sigma)$ and $(x(t), y(t)) \in \mathcal{W}^{2m} + \sigma$, we have the following estimates provided ε is small enough.

$$|I(t) - I(0)|_2 \leq \frac{8\sqrt{n-1}M_+}{M_-^2} \varepsilon^{1/2n} |\nabla H_0|_\infty,$$

for

$$|t| \leq \frac{1}{|\nabla H_0|_\infty} \exp \left(\left(\frac{M_-}{M_+} \right)^2 \frac{\rho_1}{8\sqrt{n-1}\varepsilon^{1/2n}} \right),$$

where $\rho = \rho_1 + 2\rho_2 + \rho_3$ satisfies

- $\frac{6\mu}{\rho_2^n} \leq \frac{M_+^2}{M_-^3} |\nabla H_0|_\infty^2,$
- $\frac{3e^\sigma \rho_3}{n! |\nabla H_0|_\infty} \left(\frac{\rho_1 M_-^2}{\rho_3 4\sqrt{n-1}M_+} \right)^{n+1} < \frac{4\sigma}{25}.$

The norm $|\cdot|_\infty$ is taken over $I \in \mathcal{G}^n$.

This theorem gives the estimate of the stability constant C_2 in (1.2). For a given system, we need to optimize ρ_1 under the constraints in the theorem. We see that the decomposition of ρ can be qualitatively written as $\rho = \rho_1 + c_0 \mu^{1/n} + c_1 \rho_1^{1+1/n}$, where the constants $c_0 = c_0(M_\pm, |\nabla H_0|_\infty)$ and $c_1 = c_1(M_\pm, |\nabla H_0|_\infty, n, \sigma)$. Though not solved explicitly, we expect our estimate here improves the previous results [LNN, N] since the continuous averaging method gives us an improved normal form theorem (see Theorem 3.1).

A possible application of the result is to the 3-body problem in order to get long time stabilities. This direction is already pioneered in [N]. But the mass ratio of Jupiter to the sun obtained in [N] is too small to be satisfactory. On the other hand, in [FGKR], the authors construct diffusing orbits for restricted planar 3-body problem. The diffusion time there is polynomial w.r.t. $1/\varepsilon$.

The paper is organized as follows. First we give a complete statement of the main theorem and compare it with previous results in Section 2. Then we state a normal form Theorem 3.1 about averaging in a vicinity of periodic orbits in Section 3. This is the main result that we obtain using continuous averaging, which improves the corresponding one in [LNN, N]. Then we give a brief introduction to the continuous averaging method in Section 4. After that we give a proof of Theorem 3.1 in Section 5. This section is a higher dimensional generalization of the case studied in Section 4. We try to draw analogy between the two sections. With the normal form theorem, we first show local stability result of Nekhoroshev theorem in Section 6, and then global stability in Section 7. Here local stability means the stability result in a neighborhood of a periodic orbit and global stability means stability for all initial conditions. Finally, we have two appendices A and B. The first one contains some technical estimates for the continuous averaging. The second one is some basics of majorant estimates.

2. THE COMPLETE STATEMENT OF THE MAIN THEOREM AND DISCUSSIONS

We give a complete statement of the main theorem as follows.

Theorem 2.1. *Under the same assumption as Theorem 1.1, we have*

$$|I(t) - I(0)|_2 \leq \frac{8\sqrt{n-1}M_+}{M_-^2} \varepsilon^{1/2n} |\nabla H_0|_\infty,$$

$$\text{for } |t| \leq \mathcal{T} = \frac{1}{|\nabla H_0|_\infty} \exp \left(\left(\frac{M_-}{M_+} \right)^2 \frac{\rho_1}{8\sqrt{n-1}\varepsilon^{1/2n}} \right),$$

where $\rho = \rho_1 + 2\rho_2 + \rho_3$, provided the following restrictions are satisfied.

- $\frac{6\mu}{\rho_2^n} \leq \frac{M_+^2}{M_-^3} |\nabla H_0|_\infty^2,$
- $\varepsilon^{1/2n} \leq \min \left\{ \frac{\sqrt{n-1} |\nabla H_0|_\infty^2}{5n\mu M_-} (\rho_2 + \rho_3)^n, \frac{M_-^2}{4\sqrt{n-1} |\nabla^3 H_0|_\infty}, \right.$
 $\left. \frac{M_-^2}{8\sqrt{n-1}M_+} \left(\frac{\sigma}{5(\sqrt{n} + 2\sqrt{m}) |\nabla H_0|_\infty}, \frac{\rho_1}{2(n+2m)\rho_3}, \frac{5\rho_1^2}{2\sigma\rho_3} \right) \right\}.$

$$\bullet \frac{3e^\sigma \rho_3}{n! |\nabla H_0|_\infty} \left(\frac{\rho_1 M_-^2}{\rho_3 4\sqrt{n-1} M_+} \right)^{n+1} < \frac{4\sigma}{25},$$

$$\varepsilon^{1/2n} \frac{16n\sqrt{n-1} M_+}{\rho_1 M_-^2} \left(1 + \ln \frac{\rho_1^2 M_-^4 \varepsilon^{-1/n}}{32n(n-1) M_+^2 \rho_3} \right) \leq 1.$$

The norm $|\cdot|_\infty$ is taken over $I \in \mathcal{G}^n$.

The constant μ plays the same role as the constant E in [LNN, N]. It is dual to ε since only the product $\varepsilon\mu$ enters the original Hamiltonian. We need the smallness of μ to make the first bullet point in Theorem 2.1 satisfied. The same restriction is expressed in [LNN, N] by introducing a constant g . The second bullet point can be satisfied easily by taking ε small enough. To improve the stability time, we want ρ_1 to be as large as possible, but the third bullet point gives a restriction of ρ_1 so that we need to optimize among ρ_1, ρ_2, ρ_3 . This restriction appears due to the finiteness of the width of analyticity of the action variables I and degenerate variables x, y . It shows up in a different form in [LNN] as item (ii) of Theorem 2.1, where the choice of R there can be as small as $\varepsilon^{1/2n}$. We will give more discussions in Remark 6.1 and 7.1. We will see from the following Theorem 3.1 that our normal form theorem obtained from continuous averaging improves that obtained from the iteration method. Therefore we see we also get improved C_2 here even though ρ_1 is not expressed explicitly.

3. NORMAL FORM

Our main work in this paper is to obtain a normal form theorem using continuous averaging. Following Lochak, we do the averaging in a neighborhood of a periodic orbit.

Definition 3.1. We define $\omega^* = (p_1, p_2, \dots, p_n)/\bar{T}$, $p_i \in \mathbb{Z}$, $\bar{T} \in \mathbb{R} \setminus \{0\}$, and $\text{g.c.d.}(p_1, p_2, \dots, p_n) = 1$. This is the frequency vector of a periodic orbit of the unperturbed Hamiltonian H_0 . The period T of this vector is, $T = 2\pi\bar{T}$.

Integer vectors k with $\langle \omega^*, k \rangle \neq 0$ give us

$$(3.1) \quad |\langle k, \omega^* \rangle| = |k \cdot (p_1, p_2, \dots, p_n)|/\bar{T} \geq 1/\bar{T}$$

After a proper translation in the space of action variables, wlog, we assume

$$\omega(0) := \frac{\partial H_0}{\partial I}(0) = \omega^*.$$

We can split the Hamiltonian (1.1) into four parts

$$(3.2) \quad H(I, \theta, x, y) = \langle \omega^*, I \rangle + G(I) + \varepsilon \bar{H}(I, \theta, x, y) + \varepsilon \tilde{H}(I, \theta, x, y).$$

where each of the terms is given in the next definition.

Definition 3.2. We use the Taylor expansion of H_0 to split it as

$$H_0(I) = \langle \omega^*, I \rangle + G(I),$$

where $G(I)$ contains the higher order terms. For H_1 part, we use the Fourier expansion $H_1 = \sum_{k \in \mathbb{Z}^n} H^k(I, x, y) e^{i\langle k, \theta \rangle}$ to write

$$\varepsilon H_1(I, \theta, x, y) = \varepsilon \bar{H}(I, \theta, x, y) + \varepsilon \tilde{H}(I, \theta, x, y),$$

where

$$\begin{aligned} \varepsilon \bar{H} &:= \varepsilon \sum_{\langle k, \omega^* \rangle = 0} H^k e^{i\langle k, \theta \rangle}, \quad \text{the resonant part,} \\ \varepsilon \tilde{H} &:= \varepsilon \sum_{\langle k, \omega^* \rangle \neq 0} H^k e^{i\langle k, \theta \rangle}, \quad \text{the nonresonant part.} \end{aligned}$$

The Hamiltonian equations can be written as

$$\begin{aligned} \dot{I} &= -\varepsilon \frac{\partial \bar{H}}{\partial \theta}(I, \theta, x, y) - \varepsilon \frac{\partial \tilde{H}}{\partial \theta}(I, \theta, x, y), \\ (3.3) \quad \dot{\theta} &= \omega^* + \nabla G(I) + \varepsilon \frac{\partial \bar{H}}{\partial I}(I, \theta, x, y) + \varepsilon \frac{\partial \tilde{H}}{\partial I}(I, \theta, x, y), \\ \dot{x} &= -\varepsilon \frac{\partial H_1}{\partial y}, \quad \dot{y} = \varepsilon \frac{\partial H_1}{\partial x}. \end{aligned}$$

Theorem 3.1. Suppose $|I|_2, |x|_2, |y|_2 \leq \mathcal{R}$, $(I, \theta, x, y) \in \mathcal{D}(\rho, \sigma)$ for some $\mathcal{R}, \rho, \sigma > 0$. Then there exists $\varepsilon_0 > 0$, such that for any $0 < \varepsilon < \varepsilon_0$, there exist $\rho_1, \rho_2, \rho_3 > 0$ such that $\rho_1 + 2\rho_2 + \rho_3 = \rho$ and a symplectic change of variables, $(I, \theta, x, y) \rightarrow (I', \theta', x', y')$, for $|I'|_2 \leq \mathcal{R}$, $(I', \theta', x', y') \in \mathcal{D}(\rho_2, 4\sigma/5)$, which sends the Hamiltonian (3.2) to the following normal form:

$$H = \langle \omega^*, I' \rangle + G(I') + \varepsilon \bar{\Psi}(I', \theta', x', y') + \varepsilon \tilde{\Psi}(I', \theta', x', y').$$

with the nonresonant part (see Definition 3.2)

$$\left\| \tilde{\Psi}(I', \theta', x, y) \right\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n} \exp\left(-\frac{2\pi\rho_1}{M_+\mathcal{R}T}\right),$$

the resonant part $\left\| \bar{\Psi} \right\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n}$, and the change of variables

$$|(I', \theta', x', y') - (I, \theta, x, y)|_\infty \leq \frac{5\varepsilon\mu T}{2\pi(\rho_2 + \rho_3)^n}.$$

The $\varepsilon_0, \mathcal{R}, K = \frac{2\pi\rho_1}{\rho_3 M_+ \mathcal{R} T}$ and ρ_1, ρ_2, ρ_3 satisfy the following restrictions,

- $5(\sqrt{n} + 2\sqrt{m})\mathcal{R} < \sigma$,
- $2(n + 2m) \leq K, \quad 2\sigma/(5\rho_1) \leq K$,

$$\bullet \frac{3}{\pi} e^{\sigma} \varepsilon \mu T \frac{(2K)^{n-1} K^2 \rho_3}{n!} \left(1 + \frac{2n}{\rho_3 K} \left(1 + \ln \frac{2K^2 \rho_3}{n} \right) \right) \leq \frac{4\sigma}{25}.$$

Moreover, ρ_1 can be arbitrarily close to ρ if ε is sufficiently small (see Remark 6.1).

The exponential smallness obtained here improves that of [LNN, N]. Continuous averaging enables us to get rid of some extraneous numerical factors that worsen the estimates. Moreover, our method has an advantage, that is we do not need to do a preliminary transform which is necessary in [LNN, LN]. The proof of this result is contained in Section 5.

4. A BRIEF INTRODUCTION TO THE CONTINUOUS AVERAGING

In this section, we give an introduction to the continuous averaging method. Please see the chapter 5 of [TZ] for more details. We try to explain the key points of the method that will be used in our later proof.

4.1. Derivation of the continuous averaging equation. We write the Hamiltonian (1.1) as $H(z)$, $z = (I, \theta, x, y)$. Suppose we have a symplectic change of variables $z = z(Z, \delta)$ depending on parameter δ , where $Z(\delta)$ denotes the new variables. Then we have

$$H(z) = H(z(Z, \delta)) := H(Z, \delta).$$

$$(4.1) \quad \frac{\partial H}{\partial Z} \frac{dZ}{d\delta} + \frac{\partial H}{\partial \delta} = 0.$$

If we choose the flow of $\frac{dZ}{d\delta}$ to be the Hamiltonian flow generated by a Hamiltonian isotopy $F(Z, \delta)$, i.e.

$$(4.2) \quad \frac{dZ}{d\delta} = J dF(Z, \delta), \quad J = \begin{bmatrix} 0 & -\text{Id} \\ \text{Id} & 0 \end{bmatrix},$$

with initial value $Z|_{\delta=0} = z$, then the change of variables is symplectic and we get

$$(4.3) \quad H_\delta = -\{H, F\}_Z = -\{H, F\}_z,$$

where the subscript δ means partial derivative. The last equality follows from the fact that the Poisson bracket is invariant under symplectic transformations. In the following, we only work with the variables z .

To simplify our discussion, we consider a special case of (1.1) with $m = n = 1$. A further simplification is to consider only time-periodic nonautonomous systems.

This is equivalent to requiring that $H_0(I) = I$ in equation (1.1) and $H_1(x, y, \theta)$ independent of I . From equation (4.3), we have:

$$(4.4) \quad H_\delta = -F_\theta - \{H, F\}_{(x,y)},$$

where $\{\cdot, \cdot\}_{(x,y)}$ stands for the x, y part of the Poisson bracket.

Our goal is to show that

if we choose a suitable Hamiltonian isotopy F and extend δ as large as possible, the dependence of H on θ can be killed to be exponentially small, i.e. $O(e^{-c/\varepsilon})$ for some constant c .

Suppose $H(z, \delta)$ has Fourier expansion

$$(4.5) \quad H(z, \delta) = I + \varepsilon \langle H_1 \rangle + \varepsilon \sum_{k \in \mathbb{Z} \setminus \{0\}} H^k(x, y, \delta) e^{ik\theta},$$

where $\varepsilon \langle H_1 \rangle$ means the zeroth Fourier coefficient of H_1 .

We choose the Hamiltonian isotopy F as the ‘‘Hilbert transform’’:

$$(4.6) \quad F(z, \delta) = - \sum_{k \in \mathbb{Z} \setminus \{0\}} i\varepsilon \sigma_k H^k(x, y, \delta) e^{ik\theta}, \quad \sigma_k = \text{sgn}(k).$$

Now equation (4.4) has the form in terms of Fourier coefficients:

$$(4.7) \quad H_\delta^k = -|k|H^k + i\varepsilon \sigma_k \{\langle H_1 \rangle, H^k\}_{(x,y)} + i\varepsilon \sum_{l+m=k} \sigma_m \{H^l, H^m\}_{(x,y)}, \quad k \in \mathbb{Z} \setminus \{0\}.$$

We show this F is the good choice that makes the dependence on θ decrease exponentially.

4.2. The choice of the Hamiltonian isotopy F .

4.2.1. Heuristic argument. Following [TZ], we explain here the heuristic ideas that justify this choice of F . If we set $\varepsilon = 0$ in (4.7), we get

$$H_\delta^k = -|k|H^k,$$

whose solution tends to zero as $\delta \rightarrow \infty$. If we neglect the third term in the RHS of (4.7), we have

$$H_\delta^k = -|k|H^k - i\varepsilon \sigma_k \{H^k, \langle H_1 \rangle\}_{(x,y)}.$$

It has an exact solution of the form

$$(4.8) \quad H^k(I, x, y, \delta) = e^{-|k|\delta} H^k(I, x, y, 0) \circ g^{-i\varepsilon \sigma_k \delta},$$

where g means the Hamiltonian flow generated by the Hamiltonian $\langle H_1 \rangle$. Notice the imaginary unit i here. It tells us that the flow is considered with purely imaginary time. As δ increases, the complex width of analyticity is lost gradually. So formula (4.8) has sense only if we take $\varepsilon \delta < \rho$, where ρ is the width of analyticity in θ . This is an obstacle for the extendability of δ .

We see from the heuristic argument that this choice of F gives us the exponential decay as well as a good guess for the stopping time.

4.2.2. Comparison with the Lie method. The Lie method is used in the works [N, LN, LNN]. Before working out the detailed proof of the above heuristic argument, we explain the ‘‘Hilbert transform’’ first. In fact this choice of F is strongly related to the classical averaging theory. Let us recall what we usually do in the Lie method.

Define the linear operator of taking Lie derivative along the Hamiltonian flow generated by the Hamiltonian function \hat{F} : $\mathcal{L}_{\hat{F}}H = \{H, \hat{F}\}$.

The time-1 map of (4.1) and (4.3) is:

$$\begin{aligned} H|_{\delta=1} &= e^{\mathcal{L}_{\hat{F}}} H \\ &= H + \{H, \hat{F}\} + \frac{1}{2}\{\{H, \hat{F}\}, \hat{F}\} + \cdots \\ &= H_0 + \varepsilon H_1 + \{H_0, \hat{F}\} + h.o.t. \end{aligned}$$

In each step of iteration, we need to solve the cohomological equation

$$(4.9) \quad \varepsilon H_1 + \{H_0, \hat{F}\} = 0.$$

In fact, we are only able to solve

$$(4.10) \quad \varepsilon H_1 - \varepsilon \langle H_1 \rangle + \{H_0, \hat{F}\} = 0.$$

By comparing the Fourier coefficients, we obtain the following

$$(4.11) \quad \varepsilon H^k(I) + ik\hat{F}^k = 0 \implies \hat{F}^k = i \frac{\varepsilon H^k(I)}{k}, \quad k \neq 0.$$

Now we can explain why we choose F as the Hilbert transform of H in (4.6). We select F to inherit the most important information in \hat{F} , namely, the imaginary unit i and $\text{sgn}(k)$. Readers can check that we still get the heuristic argument above if we choose the \hat{F} whose Fourier coefficients are (4.11) to do the averaging in (4.3).

4.3. The integral equation. Now we take into account the third term in the RHS of equation (4.7). We first remove the $-|k|H^k$ term in equation (4.7) by setting $H^k = e^{-|k|\delta} u^k$ to obtain

$$(4.12) \quad u_\delta^k = -i\varepsilon\sigma_k\{u^k, \langle H_1 \rangle\}_{(x,y)} + 2i\varepsilon\sigma_m \sum_{\substack{l+m=k, \\ m<0<l}} e^{-(|l|+|m|-|k|)\delta} \{u^l, u^m\}_{(x,y)}.$$

If we define an operator $g^{is*}f := f \circ g^{is}$, where g^t is the flow generated by the Hamiltonian $-\langle H_1 \rangle$, the exact solution of the truncated equation

$$u_\delta^k = -i\varepsilon\sigma_k\{u^k, \langle H_1 \rangle\}_{(x,y)}$$

would be $g^{\varepsilon\sigma_k i\delta^*} u^k(x, y, 0)$.

Then using the variation of parameter method in ODE, we can write the exact solution to equation (4.12) in the following form.

$$\begin{aligned}
 (4.13) \quad & u^k(x, y, \delta) \\
 &= g^{\varepsilon\sigma_k i\delta^*} u^k(x, y, 0) + 2i\varepsilon\sigma_m \int_0^\delta e^{-(|l|+|m|-|k|)s} g^{\varepsilon\sigma_k i(\delta-s)*} \sum_{\substack{l+m=k, \\ m<0<l}} \{u^l, u^m\}_{(x,y)} ds \\
 &= g^{\varepsilon\sigma_k i\delta^*} u^k(x, y, 0) \\
 &+ 2i\varepsilon\sigma_m \int_0^\delta \sum_{\substack{l+m=k, \\ m<0<l}} e^{-(|l|+|m|-|k|)s} \{g^{\varepsilon\sigma_k i(\delta-s)*} u^l, g^{\varepsilon\sigma_k i(\delta-s)*} u^m\}_{(x,y)} ds.
 \end{aligned}$$

We will analyze this equation to study its solution. To do so, we need a good control of the non-homogeneous term, i.e. the second term in the RHS.

4.4. Control of the nonhomogeneous term of equation (4.13). To control the nonhomogeneous term, we use the majorant estimate. The majorant relation “ \ll ” is defined as follows.

Definition 4.1. For any two functions $f(z)$, $g(z)$, $z = (z_1, z_2, \dots, z_m)$, analytic at the point $z = 0$,

$$f(z) = \sum_{\beta} f_{\beta} z^{\beta}, \quad g(z) = \sum_{\beta} g_{\beta} z^{\beta},$$

$$\beta = (\beta_1, \beta_2, \dots, \beta_m), \quad \beta_j \geq 0, \quad z^{\beta} = z_1^{\beta_1} \cdots z_m^{\beta_m}.$$

We say that g is a majorant for f ($f \ll g$) if for any multi-index β , we have $g_{\beta} \geq |f_{\beta}|$.

The proof is first to guess a majorant assumption, then show the function in the assumption satisfies an equation that majorates the integral equation (4.13). This checks the assumption and closes up the argument.

Now we make a majorant assumption

$$(4.14) \quad g^{\varepsilon\sigma_k i\tau^*} u^k(x, y, \delta) \ll \mu e^{-|k|\rho} V(Y, \delta), \quad Y = x + y, \quad |\tau| < \delta^*,$$

where $\delta^* \sim \rho/\varepsilon$ is the maximal extension time determined by the homogeneous part of equation (4.7) in the heuristic argument. The $e^{-|k|\rho}$ characterizes the way how the Fourier coefficients decay in the case of analytic perturbation and $\mu = \|H_1\|_{\rho}$. We choose $Y = x+y$ to make it easier to calculate the derivatives since $\frac{\partial V}{\partial x} = \frac{\partial V}{\partial y} = \frac{\partial V}{\partial Y}$.

Then the integrand of equation (4.13) can be majorated by $C(s)(V_Y)^2$, where

$$C(s) = 4\varepsilon\mu \sum_{\substack{l+m=k, \\ m < 0 < l}} e^{-(|l|+|m|-|k|)(s+\rho)} \leq 4\mu\varepsilon(1+1/\rho) := C.$$

This C depends on the smoothness and magnitude of H_1 and the number of combinations $l+m=k$. The number of combinations of integers in one dimensional case is easy to estimate, but in higher dimensional case it becomes very difficult, which is the main difficulty that we need to overcome in this paper.

If we can solve the equation $V_\delta = CV_Y^2$, then equation (4.13) can be viewed as

$$(4.15) \quad e^{|k|\rho} u^k \ll V(0) + \int_0^\delta V_s ds = V(0) + V(\delta) - V(0) = V(\delta).$$

This checks the majorant assumption (4.14). In order to solve the equation $V_\delta = C(V_Y)^2$, we apply the operator $\frac{\partial}{\partial Y}$ to the equation. We get the following Burgers equation by setting $U = V_Y$.

$$U_\delta = 2CUU_Y, \quad U(Y, 0) = \frac{\sigma}{\sigma - Y}.$$

Here $U(Y, 0)$ majorates ∇u in the sense $\nabla u^k(x, y, 0) \ll Me^{-|k|\rho} U(Y, 0)$. The initial condition $\frac{\sigma}{\sigma - Y}$ is due to Lemma B.1 in Appendix B, where σ is the width of analyticity in the slow variables (x, y) .

4.5. Outcome of the continuous averaging procedure. The Burgers equation can be solved explicitly using the characteristics method in PDE. The solution is

$$(4.16) \quad U(Y, \delta) = \frac{2\sigma}{(\sigma - Y) + \sqrt{(\sigma - Y)^2 - 8\sigma C\delta}}.$$

In order to ensure $(\sigma - Y)^2 - 8\sigma C\delta \geq 0$, we obtain the maximal flow time given by the slow variables is $\delta < \frac{\sigma}{8C}$. In fact $C = O(\varepsilon)$, so combined with $\delta \leq \delta^* \sim \rho/\varepsilon$ we get the maximal flow time is $O(1/\varepsilon)$. We also notice U is always bounded provided $|Y|$ is sufficiently small, so is V . Recall that we defined

$$H^k = e^{-|k|\delta} u^k \ll e^{-|k|\delta - |k|\rho} V.$$

Therefore each Fourier coefficient H^k after the continuous averaging would be less than $e^{-\delta} = O(e^{-c/\varepsilon})$ for some constant c . Adding up all these Fourier terms, we recover the Hamiltonian after the averaging, which is of order $O(e^{-c/\varepsilon})$. This is the result proved in [Tr1, Tr2, TZ]. We will work out all the details in Section 5.5.

5. CONTINUOUS AVERAGING PROOF OF THE NORMAL FORM THEOREM 3.1

Now we prove Theorem 3.1 using the continuous averaging method. Let us go back to the setup in Section 3. Since we are looking at a motion that is very close to periodic orbit in the region of the phase space, the continuous averaging explained in the previous section could be applied. The periodic orbit corresponds to the fast angle θ in equation (4.5). The nonresonant part \tilde{H} corresponds to the θ dependence term $\sum_{k \neq 0} H^k e^{ik\theta}$ in equation (4.5). The $\langle \omega^*, I \rangle$ will produce the exponential decay in the same way as the term I in equation (4.5) did in equation (4.8). And the term $G(I)$ will generate the imaginary flow in the same way as the term $\langle H_1 \rangle$ in equation (4.5) did in equation (4.8). Finally, the term \bar{H} leads to additional difficulties.

We devote the remaining part of this section to the proof of Theorem 3.1. The proof is organized as follows.

- Set up the continuous averaging in terms of Hamiltonian and get some heuristic understanding of the averaging process in Section 5.1.
- Apply it the Hamiltonian vector field in Section 5.2.
- Following procedures in Section 4, we define the operator \mathbf{g} to write the differential equations as integral equations, then we write down the majorant equation and prove the majorant relations.
- Derive necessary estimates in the theorem from the majorant estimates in Section 5.5.

5.1. Continuous averaging for Hamiltonian (3.2). In this section, we write down the continuous averaging and get a heuristic understanding. We start with a definition. As we have seen in Section 4, in the process of continuous averaging, we have different aspects like exponential decay, imaginary flow and nonhomogeneous terms.

Definition 5.1. *We define a partition of the width of analyticity ρ ,*

$$\rho = \rho_1 + 2\rho_2 + \rho_3, \quad \rho_1, \rho_2, \rho_3 > 0,$$

and put

$$K = \frac{\rho_1}{\rho_3 M_+ \mathcal{R} \bar{T}}.$$

For $\delta > 0$, we also define the following sets to form a partition of the grid \mathbb{Z}^n .

$$D_-(\delta) = \{l_- \in \mathbb{Z}^n \mid \langle l_-, \omega^* \rangle < 0, \quad |l_-| \rho_3 + |\langle l_-, \omega^* \rangle| \delta \leq \rho_3 K\},$$

$$D_+(\delta) = \{l_+ \in \mathbb{Z}^n \mid \langle l_+, \omega^* \rangle > 0, \quad |l_+| \rho_3 + |\langle l_+, \omega^* \rangle| \delta \leq \rho_3 K\},$$

$$D_0 = \{l_0 \in \mathbb{Z}^n \mid \langle l_0, \omega^* \rangle = 0\},$$

$$D_{>}(\delta) = \mathbb{Z}^n \setminus (D_-(\delta) \cup D_+(\delta) \cup D_0).$$

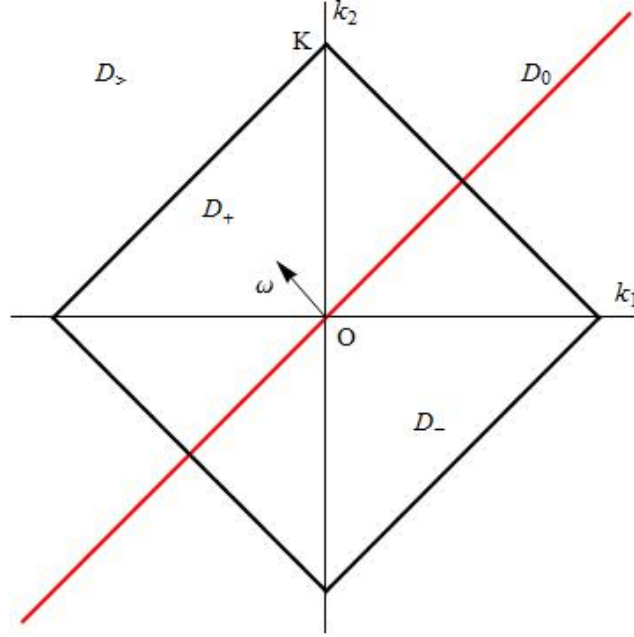


FIGURE 1. Definition of sets when $\delta = 0$. The diamond encloses integer vectors $|k| \leq K$. ω is the frequency vector ω^* . The red line is a hyperplane perpendicular to ω . The region outside the diamond is $D_{>}(\delta)$. The red line splits the region inside the diamond into D_+, D_- .

Finally, we define two functions of δ associated to the above sets.

$$\sigma_k(\delta) = \begin{cases} 1 & k \in D_+(\delta), \\ -1 & k \in D_-(\delta), \\ 0 & k \in D_0 \cup D_{>}(\delta). \end{cases}$$

$$S_k(\delta) = \int_0^\delta \sigma_k(s) ds.$$

- Remark 5.1.**
- We split the analyticity width ρ of the fast angle θ into $\rho = \rho_1 + 2\rho_2 + \rho_3$. This splitting is quite flexible. We will optimize it to make ρ_1 as large as possible in Section 6 and 7. Here ρ_1 would be used to control the imaginary flow, ρ_3 is used to do averaging, and ρ_2 is the remaining width of analyticity in angular variables after averaging. These distinctions will be made clear in the course of the proof.
 - We choose the cut-off K to make sure that if $|k| \geq K$, then the corresponding Fourier coefficient is smaller than $e^{-\rho_3 K}$, which we think to be sufficiently

small. A Fourier coefficient with $k \in D_{\pm}(\delta)$ will become smaller as δ increases. Once it is smaller than $e^{-\rho_3 K}$, the vector k enters $D_{>}(\delta)$. So $D_{\pm}(\delta)$ keeps shrinking as δ increases. We stop running the continuous averaging once $D_{\pm} = \emptyset$.

We also define

$$\delta^* := \sup\{\delta \mid D_{\pm}(\delta) \neq \emptyset\}$$

as the stopping time.

Lemma 5.1. *The stopping time δ^* satisfies*

$$\delta^* \leq K\bar{T}\rho_3.$$

Proof. From the definition of $D_{\pm}(\delta)$, we know that at time δ^* , we should have

$$|l_+|\rho_3 + |\langle l_+, \omega^* \rangle|\delta^* = \rho_3 K.$$

We know $|\langle l_+, \omega^* \rangle| \geq 1/\bar{T}$ from equation (3.1). This implies $\delta^* \leq K\bar{T}\rho_3$. \square

Now let us build our continuous averaging. This part is analogous to Section 4.1. From equation (4.3), we have

$$(5.1) \quad H_{\delta} = \{F, H\} = \{F, \langle \omega^*, I \rangle\} + \{F, G\} + \{F, \varepsilon \bar{H}\} + \{F, \varepsilon \tilde{H}\}.$$

Lemma 5.2. *If we define*

$$(5.2) \quad F := \sum_{k \in \mathbb{Z}^n} i\varepsilon \sigma_k(\delta) H^k(I, x, y, \delta) e^{i\langle k, \theta \rangle}$$

in the continuous averaging equation (5.1), where $\sigma_k(\delta)$ is defined in Definition 5.1, then depending on the properties of the Fourier mode k , we have the following three

groups of PDEs.

(5.3a)

For $k \in D_0$,

$$H_\delta^k e^{i\langle k, \theta \rangle} = \left\{ F, \varepsilon \tilde{H} \right\}^k = \sum_{l_\pm + l = k} i\sigma_{l_\pm} \left\{ H^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, \varepsilon H^l e^{i\langle l, \theta \rangle} \right\}, \quad l \in \mathbb{Z}^n \setminus D_0.$$

(5.3b)

For $k \in D_-(\delta) \cup D_+(\delta)$,

$$\begin{aligned} H_\delta^k e^{i\langle k, \theta \rangle} &= -|\langle \omega^*, k \rangle| H^k e^{i\langle k, \theta \rangle} + i\sigma_k \left\{ H^k e^{i\langle k, \theta \rangle}, G \right\} + \left\{ F, \varepsilon \bar{H} \right\}^k + \left\{ F, \varepsilon \tilde{H} \right\}^k \\ &= -|\langle \omega^*, k \rangle| H^k e^{i\langle k, \theta \rangle} + i\sigma_k \left\{ H^k e^{i\langle k, \theta \rangle}, G \right\} + \sum_{l_\pm + l = k} i\sigma_{l_\pm} \left\{ H^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, \varepsilon H^l e^{i\langle l, \theta \rangle} \right\}. \end{aligned}$$

(5.3c)

For $k \in D_>(\delta)$,

$$H_\delta^k e^{i\langle k, \theta \rangle} = \left\{ F, \varepsilon \bar{H} \right\}^k + \left\{ F, \varepsilon \tilde{H} \right\}^k = i \sum_{l_\pm + l = k} \varepsilon \sigma_{l_\pm} \left\{ H^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, H^l e^{i\langle l, \theta \rangle} \right\}.$$

We have $l \neq 0$ in the latter two cases.

Proof. From the definition of F we know the Fourier harmonics of F come only from $D_\pm(\delta)$. As a result for any $k \neq 0$, we must write $k = l_\pm + l$ for $l_\pm \in D_\pm(\delta)$ and some l . The equation (5.3b) is straightforward.

If $k \in D_0$, then $\langle k, \omega^* \rangle = 0$. We know $\langle l_\pm, \omega^* \rangle \neq 0$, then $\langle l, \omega^* \rangle \neq 0$. So in equation (5.3a), the H_0, \bar{H} terms do not appear. If $k \in D_>(\delta)$, no Fourier harmonics from $\{F, \langle I, \omega^* \rangle\} + \{F, G\}$ in equation (5.1) enter equation (5.3c). \square

Following directly from Definition 1.2, we have the lemma.

Lemma 5.3. *If we define \mathcal{R} as the confinement radius of I , i.e. $|I|_2 \leq \mathcal{R}$, $I \in G_n + \sigma \subset \mathbb{C}^n$, then we have the following estimates*

$$|G| \leq M_+ \mathcal{R}^2/2, \quad |\nabla G|_2 \leq M_+ \mathcal{R}.$$

Proof. We first notice $G(0) = \nabla G(0) = 0$. For $|G|$, we use the formula

$$G(I) = \left\langle I \int_0^1 (1-t) \nabla^2 H_0(tI) dt, I \right\rangle$$

and Definition 1.2 to get the estimate in the lemma. For $|\nabla G(I)|_2$, we use

$$\nabla G(I) = I \int_0^1 \nabla^2 G(tI) dt = I \int_0^1 \nabla^2 H_0(tI) dt.$$

□

The following lemma helps us to understand the heuristic ideas of the process of continuous averaging and Definition 5.1.

Lemma 5.4. *If we omit the $\sum_{k=l_{\pm}+l}$ terms in the RHS of equations (5.3), then equation (5.3) can be solved explicitly and the solution satisfies*

$$\begin{aligned} |H^k(I, x, y, \delta)| &\leq \mu e^{-\rho|k|}, \quad \text{for } k \in D_{>}(0) \cup D_0, \\ |H^k(I, x, y, \delta)| &\leq \mu e^{-2\rho_2|k|} e^{-\rho_3|k| - \delta|\langle \omega^*, k \rangle|}, \quad \text{for } k \in D_{\pm}(0). \end{aligned}$$

Moreover, at the stopping time δ^* we have

$$|H^k(I, x, y, \delta^*)| \leq \mu e^{-2\rho_2|k|} e^{-\rho_3 K}, \quad \text{for } k \in D_{\pm}(0),$$

where the domain of variables is $(I, x, y) \in (\mathcal{G}^n + \sigma) \times (\mathcal{W}^{2m} + \sigma)$.

Proof. It follows from Definition 1.1 that $|H^k| \leq \mu e^{-\rho|k|}$ for

$$(I, x, y) \in (\mathcal{G}^n + \sigma) \times (\mathcal{W}^{2m} + \sigma).$$

If we truncate equations (5.3), then the first and the third become $H_{\delta}^k = 0$ for $k \in D_{>}(0) \cup D_0$. So we have the corresponding estimate of $|H^k|$ stated in the lemma. However, equation (5.3c) becomes

$$H_{\delta}^k e^{i\langle k, \theta \rangle} = -|\langle \omega^*, k \rangle| H^k e^{i\langle k, \theta \rangle} + i\sigma_k \{H^k e^{i\langle k, \theta \rangle}, G\} = (-|\langle \omega^*, k \rangle| - \sigma_k \langle k, \nabla G \rangle) H^k e^{i\langle k, \theta \rangle}.$$

This equation admits an explicit solution

$$H^k(I, x, y, \delta) = e^{-|\langle \omega^*, k \rangle| \delta - \sigma_k \langle k, \nabla G \rangle \delta} H^k(I, x, y, 0).$$

So we have the estimate

$$|H^k| \leq \mu e^{-\rho|k|} e^{-|\langle \omega^*, k \rangle| \delta - \sigma_k \langle k, \nabla G \rangle \delta}, \quad k \in D_{\pm}(\delta).$$

Using Lemma 5.3, we get

$$|\sigma_k \langle k, \nabla G \rangle| \leq |\nabla G|_{\infty} \cdot |k| \leq |\nabla G|_2 \cdot |k| \leq M_+ \mathcal{R} |k|.$$

In the splitting $\rho = \rho_1 + 2\rho_2 + \rho_3$, we use ρ_1 to bound the term $\langle k, \nabla G \rangle$. Namely, we need

$$|\sigma_k \langle k, \nabla G \rangle| \delta \leq \rho_1 |k|.$$

It is enough to require that

$$(5.4) \quad \delta M_+ \mathcal{R} \leq \rho_1.$$

This also gives an upper bound for δ . We equate this upper bound with the one given in Lemma 5.1 to obtain the value of K in Definition 5.1. Now we have

$$|H^k| \leq \mu e^{-2\rho_2|k|} e^{-\rho_3|k| - |\langle \omega^*, k \rangle| \delta}, \quad k \in D_{\pm}(\delta).$$

The definition of $D_{\pm}(\delta)$ implies that once this H^k term is already $e^{-2\rho_2|k|} e^{-\rho_3 K}$, the k will enter $D_{>}(\delta)$ and not belong to $D_{\pm}(\delta)$ any more. □

5.2. Continuous averaging for a vector field. In order for the majorant estimates to be applicable to understand equations (5.3), we need to write the continuous averaging equations in terms of Hamiltonian vector field.

Definition 5.2. We introduce the following vector fields h^* , h_0 , \bar{h} , \tilde{h} corresponding to different parts $\langle I, \omega^* \rangle$, G , \bar{H} , \tilde{H} of the Hamiltonian (3.2) and f corresponding to F .

$$(5.5) \quad \begin{aligned} h^* &= (0, \omega^*, 0, 0), & h_0 &= (0, \nabla G, 0, 0), \\ \bar{h} &= J\nabla \bar{H}, & \tilde{h} &= J\nabla \tilde{H}, & f &= J\nabla F. \end{aligned}$$

We also use h^k to denote the k -th Fourier coefficients of \bar{h} and \tilde{h} . Moreover, corresponding to F in equation (5.2), we define

$$f = \sum_k i\varepsilon \sigma_k(\delta) h^k e^{i\langle k, \theta \rangle}.$$

With this definition, we can rewrite the continuous averaging equation (5.1) as follows by replacing the Poisson bracket by Lie bracket and the upper case letters H , F by the lower case letters h , f respectively.

$$h_\delta = [f, \omega^* + h_0(I) + \varepsilon \bar{h} + \varepsilon \tilde{h}]$$

Lemma 5.5. If we set $v^k = h^k e^{S_k(\delta)\langle \omega^*, k \rangle}$ (recall $S_k(\delta)$ was defined in Definition 5.1), then equations (5.3) can be rewritten in the following form in terms of Hamiltonian vector field.

(5.6a) For $k \in D_0$,

$$v_\delta^k = e^{-i\langle k, \theta \rangle} i\varepsilon \sum_{l_\pm + l = k} \sigma_{l_\pm} \left[v^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, v^l e^{i\langle l, \theta \rangle} \right] e^{-\langle \omega^*, l_\pm \rangle S_{l_\pm} + \langle \omega^*, l \rangle S_l}.$$

(5.6b) For $k \in D_-(\delta) \cup D_+(\delta)$,

$$\begin{aligned} v_\delta^k &= ie^{-i\langle k, \theta \rangle} (\sigma_k [v^k e^{i\langle k, \theta \rangle}, v_0]) \\ &+ \sum_{l_\pm + l = k} \varepsilon \sigma_{l_\pm} \left[v^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, v^l e^{i\langle l, \theta \rangle} \right] e^{-\langle \omega^*, l_\pm \rangle S_{l_\pm} + \langle \omega^*, l \rangle S_l - \langle \omega^*, k \rangle S_k}. \end{aligned}$$

(5.6c) For $k \in D_>(\delta)$,

$$v_\delta^k = ie^{-i\langle k, \theta \rangle} \sum_{l_\pm + l = k} \varepsilon \sigma_{l_\pm} \left[v^{l_\pm} e^{i\langle l_\pm, \theta \rangle}, v^l e^{i\langle l, \theta \rangle} \right] e^{-\langle \omega^*, l_\pm \rangle S_{l_\pm} + \langle \omega^*, l \rangle S_l - \langle \omega^*, k \rangle S_k}.$$

Proof. In equations (5.3), we replace the Poisson bracket by Lie bracket and the upper case letters H , F by the lower case letters h , f respectively. Then we remove the $-\langle \omega^*, k \rangle |h^k|$ in the second case as we did in Section 4.3. We set $v^k = h^k e^{S_k(\delta)\langle \omega^*, k \rangle}$. Then direct computation proves the lemma. \square

5.3. The operator \mathbf{g} and the majorant commutator. What we do next is to write the differential equations for v^k 's as integral equations. As we did in Section 4.3, we first need to define an operator g which solves the homogeneous part of equations (5.6), esp. (5.6b).

5.3.1. *The operator \mathbf{g} .*

Definition 5.3 (Section 2 of [PT]). *Let g^t be the Hamiltonian flow of the Hamiltonian vector field $h_0(I)$ generated by the Hamiltonian $G(I)$. We put $f_k = \hat{f}(I, x, y)e^{i\langle k, \theta \rangle}$ for an arbitrary analytic function \hat{f} defined on $\mathcal{D}(\rho, \sigma)$ and then define:*

$$\mathbf{g}_k^t \hat{f} = e^{-i\langle k, \theta \rangle} g_*^{-it}(f_k \circ g^{it}), \quad t \in \mathbb{R}.$$

It is shown in Section 5 and 7 of [PT] that \mathbf{g} has the following two properties.

- $v^k(I, \delta) = \mathbf{g}_k^{S_k(\delta)} v^k(I, 0)$ solves $v_\delta^k(I, \delta) = i\sigma_k(\delta)e^{-i\langle k, \theta \rangle}[v^k e^{i\langle k, \theta \rangle}, v_0]$.
 -
- $$\mathbf{g}_{k_1+k_2}^t \left(e^{-i\langle k_1+k_2, \theta \rangle} [\hat{f}_1 e^{i\langle k_1, \theta \rangle}, \hat{f}_2 e^{i\langle k_2, \theta \rangle}] \right) = e^{-i\langle k_1+k_2, \theta \rangle} [e^{i\langle k_1, \theta \rangle} \mathbf{g}_{k_1}^t \hat{f}_1, e^{i\langle k_2, \theta \rangle} \mathbf{g}_{k_2}^t \hat{f}_2].$$

With this operator \mathbf{g} , we can write differential equations (5.6) as integral equations.

Lemma 5.6. *If we denote the $\sum_{k=l_\pm+l}$ terms in equations (5.6a, 5.6b, 5.6c) by η_a, η_b, η_c respectively, then we have the following three integral equations equivalent to equations (5.6).*

$$(5.7a) \quad \text{For } k \in D_0, \quad v^k(I, \delta) = v^k(I, 0) + \varepsilon \int_0^\delta (e^{-i\langle k, \theta \rangle} \eta_a^k) ds.$$

$$(5.7b) \quad \text{For } k \in D_- \cup D_+, \quad v^k(I, \delta) = \mathbf{g}_k^{\sigma_k \delta} v^k(I, 0) + \varepsilon \int_0^\delta \mathbf{g}_k^{\sigma_k(\delta-s)} (e^{-i\langle k, \theta \rangle} \eta_b^k) ds.$$

$$(5.7c) \quad \text{For } k \in D_>(\delta), \quad v^k(I, \delta) = v^k(I, 0) + \varepsilon \int_0^\delta (e^{-i\langle k, \theta \rangle} \eta_c^k) ds.$$

Proof. The equations (5.7a) and (5.7c) are straightforward. The equation (5.7b) is an application of the first property of the operator \mathbf{g} above and the variation of parameter method in ODE. \square

5.3.2. *The majorant commutator.* We need the following majorant commutator to perform estimates.

Definition 5.4 (Section 7 of [PT]). *For any two functions $\hat{F}, \hat{G} : \mathbb{C}^{n+2m} \rightarrow \mathbb{C}$, and any two vectors $l, k \in \mathbb{Z}^n$, we define the majorant commutator:*

$$[[\hat{F}, \hat{G}]]^{l,k} = (|l| + |k|)\hat{F}\hat{G} + (n + 2m)\frac{\partial}{\partial Y}(\hat{F}\hat{G}),$$

where $Y = I + x + y$.

For this commutator, we have the following lemmas.

Lemma 5.7 (Proposition 7.1 of [PT]). *Suppose that $\hat{f}_1, \hat{f}_2 : \mathbb{C}^{n+2m} \rightarrow \mathbb{C}^{2(n+m)}$, $\hat{F}_1, \hat{F}_2 : \mathbb{C}^{n+2m} \rightarrow \mathbb{C}$, and $\hat{f}_1 \ll \hat{F}_1$, $\hat{f}_2 \ll \hat{F}_2$. Here the majorant relation $\hat{f}_i \ll \hat{F}_i$ means that \hat{F}_i majorates each component of the vector \hat{f}_i , $i = 1, 2$. Then for any $k_1, k_2 \in \mathbb{Z}^n$,*

$$e^{-i\langle k_1+k_2, \theta \rangle} [\hat{f}_1 e^{i\langle k_1, \theta \rangle}, \hat{f}_2 e^{i\langle k_2, \theta \rangle}] \ll [[\hat{F}_1, \hat{F}_2]]^{k_1, k_2}.$$

Lemma 5.8 (Proposition 7.2 of [PT]). *Suppose that $\mathbf{g}_{k_1}^\tau \hat{f}_1(y) \ll \hat{F}_1(Y)$, $\mathbf{g}_{k_2}^\tau \hat{f}_2(y) \ll \hat{F}_2(Y)$. Then*

$$\mathbf{g}_{k_1+k_2}^\tau \left(e^{-i\langle k_1+k_2, \theta \rangle} [\hat{f}_1 e^{i\langle k_1, \theta \rangle}, \hat{f}_2 e^{i\langle k_2, \theta \rangle}] \right) \ll [[\hat{F}_1, \hat{F}_2]]^{k_1, k_2}.$$

5.4. Majorant equation, the derivation and the solution.

5.4.1. *Majorant control on the initial value.* We first have majorant control on the initial value.

Lemma 5.9. *For $|\delta| \leq \delta^*$, and $\mathcal{R} < \sigma$, $k \in \mathbb{Z}^n$, we have*

$$v^k(I, x, y, 0) \ll \frac{\mu \sigma e^{-|k|\rho}}{\sigma - Y}, \quad \mathbf{g}^\delta v^k(I, x, y, 0) \ll \frac{\mu \sigma e^{-|k|(\rho_3+2\rho_2)}}{\sigma - Y}.$$

Proof. We first consider $v^k(I, x, y, 0) = h^k$. We know $|H^k|, |h^k|_\infty \leq \mu e^{-\rho|k|}$ for $(I, x, y) \in (\mathcal{G}^n + \sigma) \times (\mathcal{W}^{2m} + \sigma)$ from the definition of μ in Definition 1.1. Then we use Lemma B.1 (4) in Appendix B to obtain the majorant control of $v^k(I, x, y, 0)$.

Now we consider the effect of \mathbf{g} . The operator \mathbf{g} is defined by the Hamiltonian flow generated by the Hamiltonian $iG(I)$ in Definition 5.3. The variables I, x, y are constants of motion of this Hamiltonian flow. So \mathbf{g} only shrinks the width of analyticity in θ but has no influence on that of I, x, y . From the definition of \mathbf{g} , we see

$$|\mathbf{g}^\delta v^k(I, x, y, 0)|_\infty \leq e^{|\langle k, \nabla G \rangle| \delta} \left| v^k(I, x, y, 0) \right|_\infty.$$

We also have

$$|\langle k, \nabla G \rangle| \delta \leq M_+ \mathcal{R} \delta^* |k| \leq \rho_1 |k|$$

according to inequality (5.4). This tells us

$$|\mathbf{g}^\delta v^k(I, x, y, 0)|_\infty \leq \mu e^{-(2\rho_2 + \rho_3)|k|}, \quad \text{for } (I, x, y) \in (\mathcal{G}^n + \sigma) \times (\mathcal{W}^{2m} + \sigma).$$

Now use Lemma B.1 (4) in Appendix B again to obtain the lemma. \square

5.4.2. *Majorant equations.* The following construction is given in [PT].

Definition 5.5. Consider a continuous function $a(\delta)$. We define the functions W and $W^{|k|}$ as follows as solutions of PDEs.

$$(5.8) \quad \begin{aligned} W_\delta &= a(\delta)WW_Y, \quad W|_{\delta=0} = \frac{1}{\sigma - Y}, \\ W^{|k|}(\delta) &= W(\delta), \quad \text{for } |k| \leq K, \\ W_\delta^{|k|} &= a(\delta) \left(WW_Y^{|k|} + \frac{|k|}{K} WW^{|k|} \right), \quad W^{|k|}|_{\delta=0} = \frac{1}{\sigma - Y}, \quad \text{for } |k| > K. \end{aligned}$$

Lemma 5.10. The solutions W and $W^{|k|}$ are given explicitly by

$$(5.9) \quad \begin{aligned} W &= \frac{2}{(\sigma - Y) + \sqrt{(\sigma - Y)^2 - 4A(\delta)}}, \\ W^{|k|} &= W e^{WB(\delta)\frac{|k|}{K}}, \quad \text{for } |k| > K, \\ A(\delta) &= \int_0^\delta a(s) ds. \end{aligned}$$

The solutions are defined up to time δ^* and for Y satisfying the restrictions.

$$(5.10) \quad A(\delta^*) \leq (\sigma - (\sqrt{n} + 2\sqrt{m})\mathcal{R})^2 / 4, \quad |Y| < (\sqrt{n} + 2\sqrt{m})\mathcal{R}.$$

Proof. The fact that W and $W^{|k|}$ are exact solutions can be checked directly. To obtain the restriction for δ^* , we need to ensure $(\sigma - Y)^2 - 4A(\delta) \geq 0$ so that the square root makes sense.

We want that when $\delta = \delta^*$, we still have $|I|_2, |x|_2, |y|_2 \leq \mathcal{R}$. We know

$$|Y| \leq |I| + |x| + |y| \leq \sqrt{n}|I|_2 + \sqrt{m}|x|_2 + \sqrt{m}|y|_2 \leq (\sqrt{n} + 2\sqrt{m})\mathcal{R}.$$

\square

Remark 5.2. Let us try to understand the PDEs (5.8) heuristically. Consider

$$(5.11) \quad U_t = WU_x + VU, \quad U(x, 0) = U_0(x).$$

The way to solve it is the characteristic method. The characteristics is given by $\frac{dx}{dt} = -W$. Then we are able to write the PDE in the form: $dU/dt = VU$. Then

$U = U_0 e^{\int V dt}$. So we see that, W determines how fast we approach the intersection of characteristics, while V determines how U grows.

5.4.3. *Proof of the majorant relation:* $W, W^{|k|}$ majorate the solutions of equation (5.7). The main result of this section is summarized in the following proposition, which implies the solutions of equations (5.8) majorate that of equation (5.7).

Proposition 5.11. *For any τ such that $|\tau| + \delta \leq \delta^*$, we have the following majorant control of the solution $v^k(I, x, y, \delta)$ of equation (5.7)*

$$\mathbf{g}_k^\tau v^k(I, x, y, \delta) \ll \sigma \mu e^{-|k|(\rho_3 + 2\rho_2)} W^{|k|}(Y, \delta), \quad \text{for } k \in \mathbb{Z}^n \setminus \{0\},$$

and

$$v^k(I, x, y, \delta) \ll \sigma \mu e^{-\rho|k|} W^{|k|}(Y, \delta), \quad \text{for } k \in D_0 \cup D_{>}(0).$$

under the restriction 5.10 coming from Lemma 5.10. (The expression of $a(\delta)$ and $A(\delta)$ will be given explicitly in Lemma 5.13.) Moreover $A(\delta^*)$ is given by

$$(5.12) \quad A(\delta^*) = 6e^\sigma \sigma \varepsilon \mu \bar{T} \frac{(2K)^{n-1} K^2 \rho_3}{n!} \left(1 + \frac{2n}{\rho_3 K} \left(1 + \ln \frac{2K^2 \rho_3}{n} \right) \right).$$

Proof. We first cite Proposition A.1 in [PT].

Lemma 5.12 ([PT]). *Consider the functions $W, W^{|k|}$ defined in Definition 5.5, then the following statements are true:*

- (1) $1/(\sigma - Y) \ll W \ll W_Y$,
- (2) $W \ll W^{|k|}$,
- (3) $W_Y W^{|k|} \ll W W_Y^{|k|}$,
- (4) $W^{|k|} \ll W e^{\frac{\sigma|k|}{K}}$,
- (5) $W^{|k'|} \ll W^{|k|} e^{\sigma(|k'| - |k|)/K}$, $|k| < |k'|$.

Let us first divide equations (5.7a), (5.7b), (5.7c) by the numerators of the initial condition in Lemma 5.9, i.e. $\sigma \mu e^{-\rho|k|}$, $\sigma \mu e^{-|k|(\rho_3 + 2\rho_2)}$ and $\sigma \mu e^{-\rho|k|}$ respectively. Then we use the expression $\sigma \mu \zeta^k$ to refer to any one of $\varepsilon e^{\rho|k|} e^{-i\langle k, \theta \rangle} \eta_a^k$, $\varepsilon e^{|k|(\rho_3 + 2\rho_2)} \mathbf{g}^{\delta-s} (e^{-i\langle k, \theta \rangle} \eta_b^k)$ or $\varepsilon e^{\rho|k|} e^{-i\langle k, \theta \rangle} \eta_c^k$, (see the integrands of equations (5.7) for the definitions of η_a , η_b , η_c).

To carry out the proof, we substitute the majorant relation in Proposition 5.11 into equations (5.7) to check that equations (5.7) are majorated by equations (5.8). This is the plan proposed in [PT]. We use the majorant commutator to majorate each of the Lie bracket of ζ^k according to Lemma 5.7, 5.8,

$$\zeta^k \ll \varepsilon \sigma \mu \sum_{l_\pm + l = k} e^{-(|l_\pm| + |l| - |k|)(\rho_3 + 2\rho_2)} e^{-(S_{l_\pm} \langle \omega^*, l_\pm \rangle + S_l \langle \omega^*, l \rangle - S_k \langle \omega^*, k \rangle)} [[W, W^{|l|}]]^{l_\pm, l}.$$

Here we also use that $e^{-\rho(|l_{\pm}|+|l|-|k|)} \leq e^{-(|l_{\pm}|+|l|-|k|)(\rho_3+2\rho_2)}$ for η_a and η_c . For simplicity, we denote the exponential weight by

$$E(k, l_{\pm}, l, \delta) := e^{-(|l_{\pm}|+|l|-|k|)(\rho_3+2\rho_2)} e^{-(S_{l_{\pm}}\langle\omega^*, l_{\pm}\rangle + S_l\langle\omega^*, l\rangle - S_k\langle\omega^*, k\rangle)}.$$

Applying the definition of the majorant commutator (Definition 5.4), we get

$$\zeta^k \ll \varepsilon \sigma \mu \sum_{l_{\pm}+l=k} E(k, l_{\pm}, l, \delta) \left((|l_{\pm}| + |l|) WW^{|l|} + 2(n+2m) WW_Y^{|l|} \right).$$

Here we use Lemma 5.12(3). This gives $(WW^{|l|})_Y \ll 2WW_Y^{|l|}$. We introduce the notations

$$\begin{aligned} \Sigma_{\pm} &:= \sum_{l_{+}+l_{-}=k} E(k, l_{+}, l_{-}, \delta), \\ \Sigma_{>} &:= \sum_{l_{\pm}+l_{>}=k} E(k, l_{\pm}, l_{>}, \delta), \\ \Sigma_0 &:= \sum_{l_{\pm}+l_0=k} E(k, l_{\pm}, l_0, \delta), \end{aligned} \tag{5.13}$$

to obtain

$$\begin{aligned} \zeta^k &\ll 2\varepsilon \sigma \mu \Sigma_{\pm} \left((|l_{+}| + |l_{-}|) W^2 + (n+2m)(W^2)_Y \right) \\ &+ \varepsilon \sigma \mu \Sigma_{>} \left((|l_{\pm}| + |l_{>}|) WW^{|l_{>}|} + 2(n+2m) WW_Y^{|l_{>}|} \right) \\ &+ \varepsilon \sigma \mu \Sigma_0 \left((|l_{\pm}| + |l_0|) WW^{|l_0|} + 2(n+2m) WW_Y^{|l_0|} \right). \end{aligned}$$

The second term in the RHS is the most complicated one. We only consider this term. The other two terms are done similarly.

$$\begin{aligned} &\varepsilon \sigma \mu \Sigma_{>} \left((|l_{\pm}| + |l_{>}|) WW^{|l_{>}|} + 2(n+2m) WW_Y^{|l_{>}|} \right) \\ &\ll \varepsilon \sigma \mu \Sigma_{>} \left((2K + |k|) WW^{|k|+K} + 2(n+2m) WW_Y^{|k|+K} \right) \\ &\ll \varepsilon \sigma \mu \Sigma_{>} \left(|k| WW^{|k|+K} + 2(K+n+2m) WW_Y^{|k|+K} \right) \\ &\ll 3Ke^{\sigma} \varepsilon \sigma \mu \Sigma_{>} \left(\frac{|k|}{K} WW^{|k|} + WW_Y^{|k|} \right). \end{aligned} \tag{5.14}$$

Here $|l_{>}| \leq K + |k|$, because $l_{>} = k - l_{\pm}$, $|l_{\pm}| \leq K$. We used Lemma 5.12(5) to decrease the exponent of $W^{|l|}$. We also imposed a mild restriction:

$$2(n+2m) \leq K. \tag{5.15}$$

If $|k| \geq K$, we get the majorant equation for the $W^{|k|}$ part in equation (5.8).

If $|k| \leq K$, using Lemma 5.12(1) and $W^{|k|} = W$, we replace the last “ \ll ” in (5.14)

by

$$\ll 6Ke^\sigma \varepsilon \sigma \mu \Sigma_{>} WW_Y.$$

This is the majorant equation for W in equation (5.8).

For Σ_{\pm} and Σ_0 , we get the same majorant estimate with $\Sigma_{>}$ replaced by Σ_{\pm} and Σ_0 .

Now the problem is to find $a(\delta)$ to give bound for $6Ke^\sigma \varepsilon \sigma \mu (2\Sigma_{\pm} + \Sigma_0 + \Sigma_{>})$. We need to do some careful analysis for this and the result is summarized in the following lemma.

Lemma 5.13. *We have the following upper bound for $(2\Sigma_{\pm} + \Sigma_0 + \Sigma_{>})$, where Σ_{\pm} , Σ_0 , $\Sigma_{>}$ are defined in (5.13),*

$$(2\Sigma_{\pm} + \Sigma_0 + \Sigma_{>})(\delta) \leq \frac{(2K - 2\delta/\bar{T}\rho_2)^{n-1}}{(n-1)!} + \begin{cases} \frac{2(2K)^n}{n!} & \text{if } \delta \leq \frac{\bar{T}n}{2K}, \\ \frac{2(2K)^{n-1}\bar{T}}{(n-1)!\delta} & \text{if } \delta \geq \frac{\bar{T}n}{2K}. \end{cases}$$

If we define $a(\delta) = 6Ke^\sigma \varepsilon \sigma \mu (2\Sigma_{\pm} + \Sigma_0 + \Sigma_{>})$, then

$$A(\delta^*) \leq A(K\bar{T}\rho_3) = \int_0^{K\bar{T}\rho_3} a(s) ds$$

is equation (5.12) in Proposition 5.11.

The proof of this lemma is given in Appendix A.

This lemma gives the restriction (5.12) in Proposition 5.11. What we have shown is that each integrand of equations (5.7) has majorant estimate

$$e^{-|k|(\rho_3+2\rho_2)} \mu \zeta^k \leq W_\delta^{|k|},$$

where $W^{|k|}$ satisfies equations (5.8). Combined with the majorant control on initial condition in Lemma 5.9, this implies the LHS of equations (5.7) is majorated by $W^{|k|}$. Now the proof of the proposition is complete. \square

5.5. The system after the averaging. The continuous averaging gives us the following information about the Hamiltonian vector fields.

Lemma 5.14. *At time $\delta = \delta^*$, we have*

$$h^k \ll \sigma \mu e^{-|k|\rho} W^{|k|}, \quad \text{for } k \in D_0 \cup D_{>}(0),$$

$$h^k \ll \sigma \mu e^{-2\rho_2|k|} e^{-\rho_3 K} W^{|k|}, \quad \text{for } k \in D_{\pm}(0).$$

Proof. Recall in Lemma 5.5, we set $v^k = h^k e^{S_k(\delta)\langle\omega^*, k\rangle}$. Using the definition of $S_k(\delta)$ in Definition 5.1, we get $v_k = h_k$ for $k \in D_0 \cup D_{>}(0)$. Then Proposition 5.11 applies to such k 's. For $k \in D_{\pm}(0)$, we must have $\rho_3|k| + S_k(\delta^*)\langle\omega^*, k\rangle = \rho_3 K$ according to the definition $D_{\pm}(\delta)$. Then apply Proposition 5.11 to this case. \square

5.5.1. *The estimate of the normal form.* Now we use the information that we have obtained to prove Theorem 3.1. Let us define the change of variables

$$(I, \theta, x, y)(0) \rightarrow (I, \theta, x, y)(\delta^*) := (I', \theta', x', y')$$

obtained by the continuous averaging at the stopping time $\delta = \delta^*$. Then, the Hamiltonian (3.2) in these new variables is of the form

$$H'(I', \theta', x', y') = H_0(I') + \varepsilon \bar{\Psi}(I', \theta', x', y') + \varepsilon \tilde{\Psi}(I', \theta', x', y'),$$

where $\bar{\Psi}$ is the resonant term and $\tilde{\Psi}$ is the nonresonant term as defined in Definition 3.2. The following lemma gives estimates for the functions $\bar{\Psi}$ and $\tilde{\Psi}$.

Lemma 5.15. *Suppose $K \geq 2\sigma/(5\rho_1)$, $K \geq 2(n+2m)$, $5(\sqrt{n}+2\sqrt{m})\mathcal{R} \leq \sigma$ and $A(\delta^*) < 4\sigma^2/25$. Denote after the averaging, $\bar{H} \rightarrow \bar{H}(\delta^*) := \bar{\Psi}$ and $\tilde{H} \rightarrow \tilde{H}(\delta^*) := \tilde{\Psi}$, then*

$$\left\| \tilde{\Psi} \right\|_{\rho_2}, \left\| \frac{\partial \tilde{\Psi}}{\partial \theta} \right\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n} e^{-\frac{\rho_1}{M_+ \mathcal{R} T}}, \quad \left\| \bar{\Psi} \right\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n}, \quad \left\| \bar{\Psi} + \tilde{\Psi} \right\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n},$$

where $(I, x, y) \in (\mathcal{G}^n + 4\sigma/5) \times (\mathcal{W}^{2m} + 4\sigma/5)$ after the averaging.

Proof. Notice the I component of h^k is ikH^k . So Lemma 5.14 implies

$$\begin{aligned} \left| k\Psi^k \right|_{\infty}, \left| \Psi^k \right| &\leq \sigma\mu e^{-\rho|k|} W^{|k|}, \text{ for } k \in D_0 \cup D_{>}(0), \\ \left| \Psi^k \right| &\leq \sigma\mu e^{-2\rho_2|k|} e^{-\rho_3 K} W, \text{ for } k \in D_{\pm}(0), \end{aligned}$$

where Ψ^k 's are Fourier coefficients of $\bar{\Psi}$ and $\tilde{\Psi}$. First we get upper bound for W . From Lemma 5.10, for $|Y| \leq (\sqrt{n}+2\sqrt{m})\mathcal{R}$ we have

$$W \leq \frac{2}{\sigma - (\sqrt{n}+2\sqrt{m})\mathcal{R}} \leq 5/(2\sigma)$$

provided $5(\sqrt{n}+2\sqrt{m})\mathcal{R} \leq \sigma$. (We will see in Section 7 that the confinement radius $\mathcal{R} = o(1)$ as $\varepsilon \rightarrow 0$, so this condition is easy to satisfy.) The remaining width of analyticity for (I', x', y') becomes $4\sigma/5$. We can also replace the condition $A(\delta^*) < (\sigma - (\sqrt{n}+2\sqrt{m})\mathcal{R})^2/4$ by

$$A(\delta^*) < \frac{4\sigma^2}{25}.$$

The factor $e^{-\rho_2|k|}$ is absorbed in the $\|\cdot\|_{\rho_2}$ norm. For $k \in D_{\pm}(0)$, we are left with $\mu e^{-(\rho_3+\rho_2)K}W$. For $k \in D_{>}(0)$, we have $W^{|k|} = W e^{WA(\delta^*)\frac{|k|}{K}}$ according to (5.8). We only need to ensure $WA(\delta^*)/K \leq \rho_1$ so that $e^{WA(\delta^*)|k|/K} \leq e^{\rho_1|k|}$. This is $\sigma/K \leq 5\rho_1/2$ because $A(\delta^*) \leq 4\sigma^2/25$ and $W \leq 5/(2\sigma)$. (We will also see in Section 7 that $K \rightarrow \infty$ as $\varepsilon \rightarrow 0$, so this condition is also easy to satisfy.)

$$\begin{aligned} \|\tilde{\Psi}\|_{\rho_2} &\leq \sigma\mu W \left(e^{-\rho_3K} \sum_{|k| \leq K} e^{-\rho_2|k|} + \sum_{|k| > K} e^{-(\rho_3+\rho_2)|k|} \right) \\ &\leq \sigma\mu W e^{-\rho_3K} \left(\frac{1}{\rho_2^n} + \frac{e^{-\rho_2K}}{(\rho_3+\rho_2)^n} \right) \leq 2\sigma\mu W e^{-\rho_3K} / \rho_2^n \leq \frac{5\mu}{\rho_2^n} e^{-\rho_3K}. \end{aligned}$$

Recalling the definition of $K = \frac{\rho_1}{\rho_3 M_+ \mathcal{RT}}$ in Definition 5.1, we find

$$\|\tilde{\Psi}\|_{\rho_2} \leq \frac{5\mu}{\rho_2^n} \exp\left(-\frac{\rho_1}{M_+ \mathcal{RT}}\right).$$

Similarly, we have the estimates for $\bar{\Psi}$, $\bar{\Psi} + \tilde{\Psi}$, $\frac{\partial \tilde{\Psi}}{\partial \theta}$. □

5.5.2. The deviation of action variables in the real domain.

Lemma 5.16. *Under the same hypothesis as Lemma 5.15, after the averaging the total deviation of the variables is*

$$|(I', \theta', x', y') - (I, \theta, x, y)|_{\infty} \leq \frac{5\varepsilon\mu T}{2\pi(\rho_3 + \rho_2)^n}.$$

Here the norm $|\cdot|_{\infty}$ is taken in the real domain.

Proof. For simplicity, we consider only the I component. The other components are similar. From equation (4.2), we have

$$\frac{dI}{d\delta} = \{I, F\} = \frac{\partial F}{\partial \theta},$$

where the RHS is a real function. Then

$$(5.16) \quad I' - I = \int_0^{\delta^*} \frac{\partial F}{\partial \theta} d\delta.$$

We have

$$\frac{\partial F}{\partial \theta} = -\varepsilon \sum_{|k| \leq K} \sigma_k(\delta) k H^k e^{i\langle k, \theta \rangle},$$

since $F = \varepsilon \sum_{|k| \leq K} i\sigma_k(\delta) H^k e^{i\langle k, \theta \rangle}$ defined in Lemma 5.2. We also have

$$\sigma_k(\delta) k H^k \ll \sigma\mu e^{-(2\rho_2+\rho_3)|k|} e^{-\delta|\langle k, \omega^* \rangle|} W(Y, \delta).$$

Hence we have the estimate (recall $2\pi\bar{T} = T$.)

$$\left\| \frac{\partial F}{\partial \theta}(\delta) \right\|_{\rho_2} \leq \varepsilon \sigma \mu \sum_{|k| \leq K} e^{-(\rho_3 + \rho_2)|k|} e^{-\delta|\langle k, \omega^* \rangle|} W \leq \frac{5\varepsilon\mu}{(\rho_3 + \rho_2)^n} e^{-\delta/\bar{T}}.$$

$$\left\| \int_0^{\delta^*} \frac{\partial F}{\partial \theta}(\delta) d\delta \right\|_{\rho_2} \leq \int_0^{\delta^*} \left\| \frac{\partial F}{\partial \theta}(\delta) \right\|_{\rho_2} d\delta \leq \frac{5\varepsilon\mu}{(\rho_3 + \rho_2)^n} \int_0^{\delta^*} e^{-\delta/\bar{T}} d\delta \leq \frac{5\varepsilon\mu\bar{T}}{(\rho_3 + \rho_2)^n}.$$

□

Proof of Theorem 3.1. Lemma 5.15 and 5.16 complete the proof of the theorem. Notice the conditions of the theorem coincide with that of Lemma 5.15 and 5.16, where the last condition in the theorem is exactly $A(\delta^*) < 4\sigma^2/25$. Lemma 5.15 gives the estimate of $\tilde{\Psi}$ and Lemma 5.16 gives the estimates for the deviation of the variables. □

6. LOCAL STABILITY: STABILITY IN A VICINITY OF A GIVEN PERIODIC ORBIT

In this section, we derive stability result using the normal form Theorem 3.1. Recall in Section 3, we have set $\omega^* = \frac{\partial H_0}{\partial I}(0)$ as the frequency vector of the periodic orbit that we are considering. We consider initial condition $I(0)$ such that $|I(0)|_2 \leq r$.

Theorem 6.1. (*Local stability*) *If the conditions of Theorem 3.1 and the following conditions are satisfied,*

- $\frac{6\varepsilon\mu}{\rho_2^n} M_- \leq M_+^2 r^2,$
- $\frac{5n\varepsilon\mu\bar{T}}{(\rho_2 + \rho_3)^n} \leq r,$

then we have stability result: if the initial conditions $|I(0)|_2 \leq r$, then one has $|I(t)|_2 \leq \mathcal{R} := 8 \frac{rM_+}{M_-}$, for all time $|t| \leq \mathcal{T} := \frac{1}{|\omega^|} e^{\frac{\rho_1}{M_+ \mathcal{R} \bar{T}}}$.*

Proof. The integrable part has the Taylor expansion around the point $I'(0)$,

$$H_0(I'(t)) - H_0(I'(0)) = \langle \omega(I'(0)), I'(t) - I'(0) \rangle$$

$$+ \left\langle \int_0^1 (1-s) \nabla^2 H_0(sI'(t) + (1-s)I'(0)) ds (I'(t) - I'(0)), I'(t) - I'(0) \right\rangle.$$

We obtain the following inequality using Definition 1.2

(6.1)

$$\frac{1}{2} M_- |I'(t) - I'(0)|_2^2 \leq |H_0(I'(t)) - H_0(I'(0))| + |\langle \omega(I'(0)), I'(t) - I'(0) \rangle|.$$

We use the energy conservation and Lemma 5.15 for the first term of the RHS to get

$$|H_0(I'(t)) - H_0(I'(0))| \leq \varepsilon \left(\left\| \left(\bar{\Psi} + \tilde{\Psi} \right) (I'(0)) \right\|_{\rho_2} + \left\| \left(\bar{\Psi} + \tilde{\Psi} \right) (I'(t)) \right\|_{\rho_2} \right) \leq \frac{10\varepsilon\mu}{\rho_2^n}.$$

For the second term in the RHS of inequality (6.1), we have

$$|\langle \omega(I'(0)), I'(t) - I'(0) \rangle| \leq |\langle \omega^*, I'(t) - I'(0) \rangle| + |\langle \omega(I'(0)) - \omega^*, I'(t) - I'(0) \rangle|,$$

For the first term in the RHS, we use the Hamiltonian equation, Lemma 5.15 and the fact that $\left\langle \omega^*, \frac{\partial \bar{\Psi}}{\partial \theta} \right\rangle = 0$.

$$|\langle \omega^*, I'(t) - I'(0) \rangle| \leq |t| \left\| \left\langle \omega^*, \frac{\partial \varepsilon \tilde{\Psi}}{\partial \theta} \right\rangle \right\|_{\rho_3} \leq \mathcal{T} \frac{5\varepsilon\mu}{\rho_2^n} e^{-\frac{\rho_1}{M_+ \mathcal{R} T}} |\omega^*|, \quad \text{for } |t| \leq \mathcal{T}.$$

For the second term,

$$\begin{aligned} |\langle \omega(I'(0)) - \omega^*, I'(t) - I'(0) \rangle| &\leq M_+ (|I(0)|_2 + |I'(0) - I(0)|_2) \cdot |I'(t) - I'(0)|_2 \\ &\leq M_+ \left(r + n \frac{5\varepsilon\mu \bar{T}}{(\rho_3 + \rho_2)^n} \right) \cdot |I'(t) - I'(0)|_2, \end{aligned}$$

where we use $|I(0)|_2 \leq r$ and $|I'(0) - I(0)|_2 \leq n \frac{5\varepsilon\mu \bar{T}}{(\rho_3 + \rho_2)^n}$ following from Lemma 5.16. We have a factor n since we go from $|\cdot|_\infty$ to $|\cdot|_2$.

If we set $a = |I'(t) - I'(0)|_2$, then we get an inequality of a from (6.1):

$$\frac{M_-}{2} a^2 \leq \frac{10\varepsilon\mu}{\rho_2^n} + \mathcal{T} \frac{5\varepsilon\mu}{\rho_2^n} e^{-\frac{\rho_1}{M_+ \mathcal{R} T}} |\omega^*| + M_+ \left(r + n \frac{5\varepsilon\mu \bar{T}}{(\rho_3 + \rho_2)^n} \right) a.$$

We choose

$$(6.2) \quad \mathcal{T} \leq \frac{1}{|\omega^*|} e^{\frac{\rho_1}{M_+ \mathcal{R} T}}, \quad n \frac{5\varepsilon\mu \bar{T}}{(\rho_3 + \rho_2)^n} \leq r,$$

to obtain

$$a \leq \frac{2M_+ r + \sqrt{4M_+^2 r^2 + \frac{30\varepsilon\sigma\mu}{\sigma\rho_2^n} M_-}}{M_-}.$$

We set

$$(6.3) \quad \frac{6\varepsilon\mu}{\rho_2^n} M_- \leq M_+^2 r^2.$$

Then we have

$$a = |I'(t) - I'(0)|_2 \leq 5 \frac{M_+ r}{M_-},$$

$$|I(t) - I(0)|_2 \leq |I(t) - I'(t)|_2 + |I'(t) - I'(0)|_2 + |I'(0) - I(0)|_2 \leq 2r + \frac{5rM_+}{M_-},$$

$$|I(t)|_2 \leq |I(0)|_2 + |I(t) - I(0)|_2 \leq 3r + 5 \frac{rM_+}{M_-} \leq \frac{8rM_+}{M_-}.$$

The proof is now complete. \square

Remark 6.1. We introduce a restriction 6.3 instead of introducing a constant g as did in [LNN, N]. The two restrictions of this theorem implies ρ_2, ρ_3 can be sufficiently small if ε is. Then ρ_1 can be very close to ρ . The restrictions of Theorem 3.1 are also satisfied for ε small enough. Then we get improved stability time compared with [LNN, N].

7. GLOBAL STABILITY: STABILITY FOR ARBITRARY INITIAL DATA

In this section, we consider stability result for arbitrary initial data and give a proof of Theorem 2.1. We first prove the following lemma.

Lemma 7.1. Let us fix $(1 <) Q \in \mathbb{R}$ and assume $\frac{\sqrt{n-1}}{Q^{1/(n-1)}} < \frac{M_-^2}{4 \sup_{I \in \mathcal{G}^n} |\nabla^3 H_0|_\infty}$, then for any $I \in \mathcal{G}^n$, there exists an integer q , $1 \leq q < Q$, and a point $I^* \in \mathcal{G}^n$ such that $|I - I^*|_2 \leq \frac{\sqrt{n-1}}{M_- \bar{T} Q^{1/(n-1)}}$ and $\omega(I^*)$ is rational vector of period $\bar{T} = q/|\omega(I)|_\infty$.

Proof. First recall the Dirichlet theorem for simultaneous approximation: for any $\alpha \in \mathbb{R}^n, Q \in \mathbb{R}$, and $Q > 1$, there exists an integer q , $1 \leq q < Q$, such that $|q\alpha - \mathbb{Z}^n|_\infty \leq Q^{-1/n}$.

An improvement of the estimate can be obtained by rescaling α to $\alpha/|\alpha|_\infty$. Then apply the Dirichlet theorem to approximate the remaining $n-1$ components of α with one of whose ± 1 entries removed. We get the following:

there exists a rational vector α^* of period $\bar{T} = \frac{q}{|\alpha|_\infty}$, $q \in \mathbb{N}$, $1 < q < Q$ and $|\alpha^* - \alpha|_\infty \leq \frac{1}{\bar{T} Q^{1/(n-1)}}$ (see (the only) Proposition in [N]).

The frequency vector is $\omega(I) = \nabla H_0(I)$. Consider two points I^* and I such that $\omega(I^*)$ is as stated in the lemma and approximates $\omega(I)$ in the same way as α^* approximates α .

$$\omega(I) - \omega(I^*) = \int_0^1 \nabla \omega(tI + (1-t)I^*) dt (I - I^*).$$

Hence from Definition 1.2

$$\begin{aligned} M_- |I - I^*|_2^2 &\leq \left\langle \int_0^1 \nabla \omega(tI + (1-t)I^*) dt (I - I^*), (I - I^*) \right\rangle \\ &= \langle (I - I^*), \omega(I) - \omega(I^*) \rangle \leq |I - I^*|_2 |\omega(I) - \omega(I^*)|_2. \end{aligned}$$

This implies

$$M_-|I - I^*|_2 \leq |\omega(I) - \omega(I^*)|_2 \leq \sqrt{n-1}|\omega(I) - \omega(I^*)|_\infty \leq \frac{\sqrt{n-1}|\omega(I)|_\infty}{qQ^{1/(n-1)}}.$$

In order to make sure the point I^* can be found given I , we need to show the frequency map $\omega(I)$ can be inverted, which can be done within the ball $B\left(\omega(I^*), \frac{M_-^2}{4|\nabla^3 H_0|_\infty}\right)$

centered at $\omega(I^*)$ with radius $\frac{M_-^2}{4|\nabla^3 H_0|_\infty}$ using implicit function theorem (see [LNN]).

So we assume $\frac{\sqrt{n-1}}{Q^{1/(n-1)}} < \frac{M_-^2}{4|\nabla^3 H_0|_\infty}$. \square

Proof of Theorem 2.1. We define $r = \frac{\sqrt{n-1}}{M_- \bar{T} Q^{1/(n-1)}}$ and $\mathcal{R} = 8r \frac{M_+}{M_-} = \frac{8\sqrt{(n-1)}M_+}{M_-^2 \bar{T} Q^{\frac{1}{n-1}}}$

as in Lemma 7.1. If we set $Q = \varepsilon^{-\frac{n-1}{2n}}$, then we have

$$\mathcal{R} = \frac{8\sqrt{n-1}M_+}{M_-^2 \bar{T}} \varepsilon^{1/2n} \leq \frac{8\sqrt{n-1}M_+}{M_-^2} \varepsilon^{1/2n} |\nabla H_0|_\infty, \quad \mathcal{R} \bar{T} = \frac{8\sqrt{n-1}M_+}{M_-^2} \varepsilon^{1/2n}.$$

The stability time in Theorem 6.1 is

$$\mathcal{T} = \frac{1}{|\omega^*|} \exp\left(\frac{\rho_1}{M_+ \mathcal{R} \bar{T}}\right) \geq \frac{1}{\sup_{I \in \mathcal{G}^n} |\nabla H_0|} \exp\left(\frac{\rho_1 M_-^2}{8\sqrt{n-1}M_+^2} \varepsilon^{-\frac{1}{2n}}\right).$$

Now let us analyze the restrictions that we have. The restrictions are from Theorem 3.1, Theorem 6.1 and Lemma 7.1. The quantities $\bar{T}, r, \mathcal{R}, K$ satisfy the following:

$$1 < \bar{T}|\omega|_\infty < Q = \varepsilon^{-\frac{n-1}{2n}}, \quad K = \frac{\rho_1}{\rho_3 \mathcal{R} \bar{T}} = \frac{\rho_1 M_-^2}{\rho_3 8\sqrt{n-1}M_+} \varepsilon^{-1/2n}.$$

$$\varepsilon^{1/2}|\omega|_\infty \frac{\sqrt{n-1}}{M_-} \leq r \leq \varepsilon^{1/2n}|\omega|_\infty \frac{\sqrt{n-1}}{M_-}.$$

$$\varepsilon^{1/2}|\omega|_\infty \frac{8\sqrt{n-1}M_+}{M_-^2} \leq \mathcal{R} \leq \varepsilon^{1/2n}|\omega|_\infty \frac{8\sqrt{n-1}M_+}{M_-^2}.$$

We substitute the bounds for $\bar{T}, r, \mathcal{R}, K$ into the restrictions that we have to obtain the restriction of Theorem 2.1. The first restriction in Theorem 2.1 is from the first one of Theorem 6.1. The second in Theorem 2.1 is a collection of the first two ones of Theorem 3.1, the second of Theorem 6.1 and that of Lemma 7.1. The last in Theorem 2.1 is from the third one of Theorem 3.1. We break it into two inequalities by setting $\frac{2n}{\rho_3 K} \left(1 + \ln \frac{2K^2 \rho_3}{n}\right) \leq 1$. \square

Remark 7.1. *The first restriction in Theorem 2.1 can be satisfied by making μ smaller while ε larger. This will lead to shorter stability time. The μ here plays the same role of the factor g in $[\text{LNN}, N]$. The second restriction can always be satisfied by making ε small. The third restriction can be satisfied by making μ or ρ_1/ρ_3 small. However, since $n!$ grows very fast, for large n , this restriction is easy to satisfy.*

APPENDIX A. PROOF OF LEMMA 5.13

The proof is done in the following Claim 1,2,3, which estimates $\Sigma_{\pm}, \Sigma_{>}, \Sigma_0$ in Lemma 5.13 respectively. Before the proof of the lemma, let us first analyze the geometry of numbers involved.

A.1. The geometry of integer vectors. Let us look at the Figure 2.

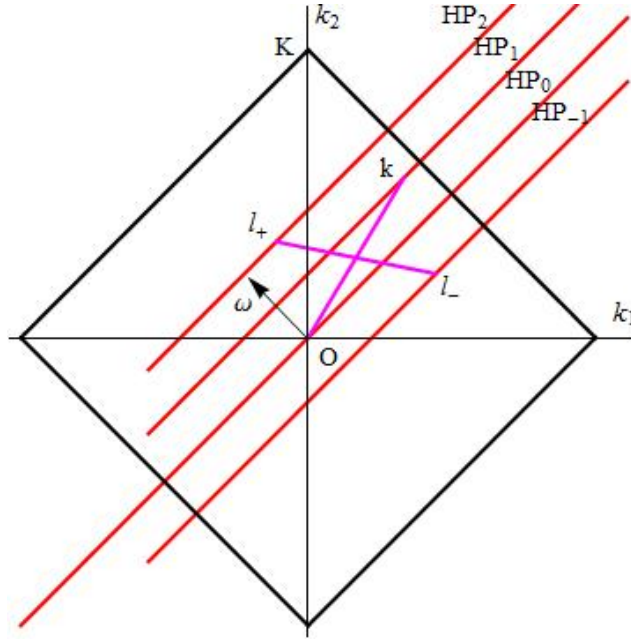


FIGURE 2. counting the number of combinations.

- *The diamond:* the diamond in the figure encloses all the vectors k with $|k| \leq K$ (in 3-dim it is an octahedron. In general it is a ball of radius K under the l^1 norm). The total number of integer vectors inclosed in the

diamond is $\frac{(2K)^n}{n!}$. Indeed, in n -dim, the diamond consists of 2^n simplices.

Each of the simplices has volume $\frac{K^n}{n!}$.

- *The hyperplane:* the small arrow indicates the rational frequency ω^* . HP_0 is a hyperplane that is perpendicular to ω^* . $\text{HP}_0 = D_0$ and the $(n-1)$ -dim volume of $\text{HP}_0 \cap \text{Diamond}$ is less than $\frac{(2K)^{n-1}}{(n-1)!}$, which is the $(n-1)$ -dim volume of an $(n-1)$ -dim Diamond. Any vector lies above HP_0 has positive inner product with ω^* , while any vector below has negative inner products. Moreover, if two vectors lie on the same hyperplane which is parallel to HP_0 , they will have the same inner product with ω^* . Let us denote $\text{HP}_d = \{k \in \mathbb{Z}^n | \langle k, \omega^* \rangle = d/\bar{T}\}$. $\text{HP}_0 \cap \text{Diamond}$ contains at most $\frac{(2K)^{n-1}}{(n-1)!}$ integer points.
- *The parallelogram:* consider the vectors l_+, l_-, k in the Figure 2. Suppose we have the relation $l_+ + l_- = k$. Then the three vectors together with the origin form a parallelogram. Suppose $\langle k, \omega^* \rangle \bar{T} = 1$, and $\langle l_+, \omega^* \rangle \bar{T} = 2$, then $\langle l_-, \omega^* \rangle \bar{T} = -1$. l_+ and l_- can move on their corresponding hyperplane, but a parallelogram is always preserved.
- *The shape of the diamond under the averaging flow:* in the definition of $D_{\pm}(\delta)$, we have the restriction $|l_{\pm}| \rho_3 + |\langle l_{\pm}, \omega^* \rangle| \delta \leq \rho_3 K$ for $l_{\pm} \in D_{\pm}(\delta)$. When $\delta = 0$, this is our diamond. When δ increases, The diamond will collapse, i.e. the integer vectors becomes fewer on HP_d .

$$(A.1) \quad |l_{\pm}| \leq K - \frac{\delta}{\rho_3} |\langle l_{\pm}, \omega^* \rangle|.$$

The rate of decreasing depends on the inner product $|\langle l_{\pm}, \omega^* \rangle|$. The farther a hyperplane HP_d is away from HP_0 (The larger the d), the faster it collapses (with volume decreasing rate $d/(\rho_3 \bar{T})$). HP_0 does not change at all. When $\delta = \delta^*$, the diamond would collapse to its intersection with HP_0 . By then we would have successfully killed all the nonresonant terms up to the desired exponential smallness $e^{-\rho_2 |k| - \rho_3 K}$. We denote the collapsed diamond at time δ by $\text{Diamond}(\delta)$.

A.2. Estimate of $\Sigma_{\pm}, \Sigma_{>}, \Sigma_0$ and the proof of Lemma 5.13. Now we obtain estimates of $\Sigma_{\pm}, \Sigma_{>}, \Sigma_0$ for fixed k .

A.2.1. Claim 1: The sum $\Sigma_{\pm}(\delta)$ defined in equation (5.13) can be estimated as follows:

$$\Sigma_{\pm}(\delta) \leq \begin{cases} \frac{(2K)^n}{2n!} & \text{if } \delta \leq \frac{\bar{T}n}{2K}, \\ \frac{(2K)^{n-1}\bar{T}}{2(n-1)!\delta} & \text{if } \delta \geq \frac{\bar{T}n}{2K}. \end{cases}$$

Proof. In the proof, the vector k is fixed. The Σ_{\pm} is defined in equation (5.13),

$$\Sigma_{\pm}(\delta) = \sum_{l_+ + l_- = k} e^{-(|l_+| + |l_-| - |k|)(\rho_3 + 2\rho_2)} e^{-(S_{l_+}\langle\omega^*, l_+\rangle + S_{l_-}\langle\omega^*, l_-\rangle - S_k\langle\omega^*, k\rangle)},$$

which can be estimated as

$$\Sigma_{\pm}(\delta) \leq \sum_{l_+ + l_- = k} e^{-(S_{l_+}\langle\omega^*, l_+\rangle + S_{l_-}\langle\omega^*, l_-\rangle - S_k\langle\omega^*, k\rangle)},$$

where we have dropped $e^{-(\rho_3 + 2\rho_2)(|l_+| + |l_-| - |k|)}$, since

$$l_+ + l_- = k \implies |l_+| + |l_-| \geq |k| \implies e^{-(\rho_3 + 2\rho_2)(|l_+| + |l_-| - |k|)} \leq 1.$$

From the relation $\langle\omega^*, l_+\rangle + \langle\omega^*, l_-\rangle = \langle\omega^*, k\rangle$, and the inequality $\delta \geq |S_k(\delta)|$, we have the following two cases depending on the sign of $\langle\omega^*, k\rangle$.

If $\langle\omega^*, k\rangle \geq 0$, then

$$\begin{aligned} S_{l_+}\langle\omega^*, l_+\rangle + S_{l_-}\langle\omega^*, l_-\rangle - S_k\langle\omega^*, k\rangle &\geq \delta(\langle\omega^*, k\rangle - \langle\omega^*, l_-\rangle) - \delta\langle\omega^*, l_-\rangle - S_k\langle\omega^*, k\rangle \\ &\geq (\delta - S_k)\langle\omega^*, k\rangle - 2\delta\langle\omega^*, l_-\rangle \geq 2\delta|\langle\omega^*, l_-\rangle|. \end{aligned}$$

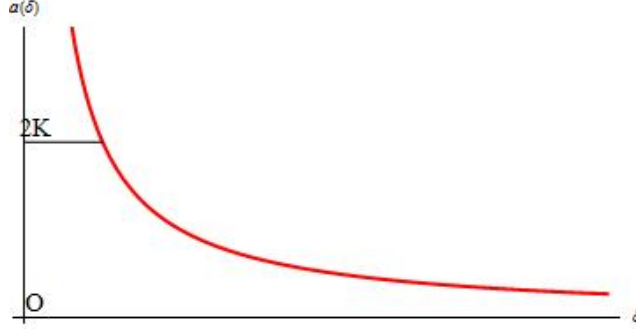
If $\langle\omega^*, k\rangle \leq 0$, then

$$\begin{aligned} S_{l_+}\langle\omega^*, l_+\rangle + S_{l_-}\langle\omega^*, l_-\rangle - S_k\langle\omega^*, k\rangle &\geq \delta\langle\omega^*, l_+\rangle - \delta(S_k\langle\omega^*, k\rangle - \langle\omega^*, l_+\rangle) - S_k\langle\omega^*, k\rangle \\ &\geq (-\delta - S_k)\langle\omega^*, k\rangle + 2\delta\langle\omega^*, l_+\rangle \geq 2\delta|\langle\omega^*, l_+\rangle|. \end{aligned}$$

Moreover, when $\delta = 0$, the number of integer vectors contained in $\text{HP}_d \cap \text{Diamond}(0)$ is no greater than $\frac{(2K)^{n-1}}{(n-1)!}$. It is zero when $|d| > K$. Since k is fixed, we can vary either l_+ or l_- . The other one will be determined uniquely. According to the analysis above, we sum over l_+ if $\langle\omega^*, k\rangle \geq 0$ while over l_- otherwise. We consider the l_+ case for instance. As δ increases, on each HP_d the number of integer vectors contained in $\text{HP}_d \cap \text{Diamond}(\delta)$ is no greater than $\frac{(2K - 2d\delta/\bar{T}\rho_3)^{n-1}}{(n-1)!}$ according to inequality (A.1). It is zero when $|d| > K$ or $K \leq d\delta/\bar{T}\rho_3$. Now we have the estimate

$$\begin{aligned} \Sigma_{\pm}(\delta) &\leq \sum_d \frac{(2k - 2d\delta/\bar{T}\rho_3)^{n-1}}{(n-1)!} e^{-2d\delta/\bar{T}} \leq \int_0^{K \wedge \frac{K\bar{T}\rho_3}{\delta}} \frac{(2K - 2d\delta/\bar{T}\rho_3)^{n-1}}{(n-1)!} e^{-2d\delta/\bar{T}} d(d) \\ &\leq \frac{(2K)^{n-1}}{(n-1)!} \int_0^\infty e^{-2d\delta/\bar{T}} d(d) \leq \frac{(2K)^{n-1}\bar{T}}{2(n-1)!\delta}. \end{aligned}$$

This estimate is poor when δ is close to zero. But in fact when $\delta = 0$, the upper bound is $\Sigma_{\pm}(0) \leq \frac{(2K)^n}{2n!}$, and the upper bound is monotonically decreasing w.r.t. δ . So we can use the bound stated in Claim 1. See figure 3. \square

FIGURE 3. upper bound of $a(\delta)$.

A.2.2. *Claim 2: The sum $\Sigma_{>}(\delta)$ defined in equation (5.13) can be estimated as follows:*

$$\Sigma_{>}(\delta) \leq \begin{cases} \frac{(2K)^n}{(2K)^{n-1}\bar{T}} & \text{if } \delta \leq \frac{\bar{T}n}{2K}, \\ \frac{n!}{(n-1)!\delta} & \text{if } \delta \geq \frac{\bar{T}n}{2K}. \end{cases}$$

Proof. It is defined in equation (5.13) that

$$\Sigma_{>}(\delta) = \sum_{l_{\pm}+l_{>}=k} e^{-(|l_{\pm}|+|l_{>}|-|k|)(\rho_3+2\rho_2)} e^{-(S_{l_{\pm}}\langle\omega^*, l_{\pm}\rangle + S_{l_{>}}\langle\omega^*, l_{>}\rangle - S_k\langle\omega^*, k\rangle)}$$

We first show

$$\Sigma_{>}(\delta) \leq \sum_{l_{\pm}+l_{>}=k} e^{-S_{l_{\pm}}\langle\omega^*, l_{\pm}\rangle}.$$

According to the definition of $l_{>} \in D_{>}(\delta)$, the Fourier term corresponding to $l_{>}$ is of the size

$$e^{-S_{l_{>}}\langle\omega^*, l_{>}\rangle - |l_{>}|\rho_3} \leq e^{-\rho_3 K}.$$

But $e^{-S_k\langle\omega^*, k\rangle - \rho_3|k|} \geq e^{-\rho_3 K}$ due to Definition 5.1. The equality is achieved only when $k \in D_{>}(\delta)$. So we know

$$e^{-S_{l_{>}}\langle\omega^*, l_{>}\rangle - |l_{>}|\rho_3 + S_k\langle\omega^*, k\rangle + \rho_3|k|} \leq 1.$$

We drop the $e^{-|l_{+}|\rho_3}$ and $e^{-(|l_{\pm}|+|l_{>}|-|k|)(2\rho_2)}$ to obtain $\Sigma_{>}(\delta) \leq \sum_{l_{\pm}+l_{>}=k} e^{-S_{l_{\pm}}\langle\omega^*, l_{\pm}\rangle}$.

This is essentially the same as the Case Σ_{\pm} discussed above. So we have the bound stated in Claim 2. \square

A.2.3. *Claim 3: The sum $\Sigma_0(\delta)$ defined in equation (5.13) can be estimated as follows:*

$$\Sigma_0(\delta) \leq \frac{(2K - 2|\langle \omega^*, k \rangle| \delta / \rho_3)^{n-1}}{(n-1)!} \leq \frac{(2K - 2\delta / \bar{T} \rho_3)^{n-1}}{(n-1)!}.$$

Proof. The term $\Sigma_0(\delta)$ turns out to be the most troublesome term. Again equation (5.13) tells us

$$\Sigma_0(\delta) = \sum_{l_{\pm} + l_0 = k} e^{-(|l_{\pm}| + |l_0| - |k|)(\rho_3 + 2\rho_2)} e^{-(S_{l_{\pm}} \langle \omega^*, l_{\pm} \rangle + S_{l_0} \langle \omega^*, l_0 \rangle - S_k \langle \omega^*, k \rangle)}$$

Because of the relation $\langle \omega^*, l_{\pm} \rangle + \langle \omega^*, l_0 \rangle = \langle \omega^*, k \rangle$ and $\langle \omega^*, l_0 \rangle = 0$, we get $\langle \omega^*, l_{\pm} \rangle = \langle \omega^*, k \rangle$ and

$$S_{l_{\pm}}(\delta) \langle \omega^*, l_{\pm} \rangle \geq S_k(\delta) \langle \omega^*, k \rangle.$$

The reason is, we know that $l_{\pm} \in D_{\pm}(\delta)$ until time δ , but we do not know where k is. This implies $|S_{l_{\pm}}(\delta)| \geq |S_k(\delta)|$. The “=” is achieved only if $k \in D_{-}(\delta) \cup D_{+}(\delta)$. So we get

$$\Sigma_0(\delta) \leq \sum_{l_{\pm} + l_0 = k} e^{-(|l_{\pm}| + |l_0| - |k|)(\rho_3 + 2\rho_2)} \leq \sum_{l_{\pm} + l_0 = k} 1.$$

Now consider Figure 2. Since $l_0 \in \text{HP}_0$, we get l_{\pm} and k must lie on the same hyperplane. So $\sum_{l_{\pm} + l_0 = k} 1$ is bounded by the number of the possible l_{\pm} ’s, which is

$$\frac{(2K - 2|\langle \omega^*, k \rangle| \delta / \rho_3)^{n-1}}{(n-1)!}.$$

This gives the Claim 3. □

Proof of Lemma 5.13. We simply add up the upper bounds for $\Sigma_0, \Sigma_{>}, \Sigma_{\pm}$ to get an upper bound for $2\Sigma_{\pm} + \Sigma_{>} + \Sigma_0$. This proves Lemma 5.13. □

APPENDIX B. ELEMENTS ON MAJORANT ESTIMATES

In this appendix, we collect some basics about the majorant relation. The materials can be found in the Chapter 5 of [TZ]. The majorant relation “ \ll ” is defined in Definition 4.1.

Lemma B.1. *The relation “ \ll ” satisfies the following properties:*

- (1) *If $f_1 \ll g_1$ and $f_2 \ll g_2$, then $f_1 + f_2 \ll g_1 + g_2$ and $f_1 f_2 \ll g_1 g_2$.*
- (2) *If $f \ll g$, then $\frac{\partial f}{\partial z_j} \ll \frac{\partial g}{\partial z_j}$ for any $j = 1, \dots, m$.*

(3) If $f(z, \lambda) \ll g(z, \lambda)$ for any value of the parameter $\lambda \in [a, b]$, then

$$\int_a^b f(z, \lambda) d\lambda \ll \int_a^b g(z, \lambda) d\lambda$$

(4) Let $|f(z)| \leq c$ in the domain $\{z = (z_1, \dots, z_m) : |z| \leq b, j = 1, \dots, m\}$.

Then $f(z) \ll c/w \ll \frac{bc}{b-z}$, where $w = b^{-m}(b - z_1) \cdots (b - z_m)$ and $z = z_1 + z_2 + \dots + z_m$.

Moreover, the majorant relation is also preserved by solving differential equations or integral equations.

Definition B.1. Consider an ODE system

$$(B.1) \quad f_\delta^k(z, \delta) = F^k(f(z, \delta), z, \delta), \quad f^k(z, 0) = \hat{f}^k(z)$$

with some known functional F^k and initial data \hat{f}^k . We call a system

$$(B.2) \quad \mathbf{f}_\delta^k(z, \delta) = \mathbf{F}^k(\mathbf{f}(z, \delta), z, \delta), \quad \mathbf{f}^k(z, 0) = \hat{\mathbf{f}}^k(z)$$

a majorant system associated with equation (B.1) if

(a) $\hat{f}^k(z) \ll \hat{\mathbf{f}}^k(z)$ for any $k \in \mathbb{Z}$, and

(b) $F^k(g(z), z, \delta) \ll \mathbf{F}^k(\mathbf{g}(z), z, \delta)$ for any $k \in \mathbb{Z}$, $\delta \geq 0$, and g, \mathbf{g} such that $g \ll \mathbf{g}$.

We have the theorem

Theorem B.1 (Chapter 5 of [TZ]). If $\mathbf{f}(z, \delta), 0 \leq \delta \leq \delta_0$ is a solution of the majorant system (B.2) associated with (B.1), then the system (B.1) has a solution and $f^k(z, \delta) \ll \mathbf{f}^k(z, \delta)$ for any $\delta \in [0, \delta_0]$, $k \in \mathbb{Z}$. The same is true if we rewrite systems (B.1) (B.2) in the integral form:

$$(B.3) \quad \begin{aligned} f^k(z, \delta) &= \hat{f}^k(z) + \int_0^\delta F^k(f(z, s), z, s) ds, \\ \mathbf{f}^k(z, \delta) &= \hat{\mathbf{f}}^k(z) + \int_0^\delta \mathbf{F}^k(\mathbf{f}(z, s), z, s) ds, \end{aligned}$$

With this theorem, we treat δ as a parameter instead of a variable. So we do not need to do the Taylor expansion w.r.t. δ .

ACKNOWLEDGEMENT

The author would like to thank Prof. Vadim Kaloshin to provide the idea of applying continuous averaging to the Nekhoroshev theorem. He also would like to thank Prof. Treschev, Neishtadt, Dr. M. Guardia and A. Bounemoura for carefully reading the manuscript and giving many valuable suggestions. The last version of the paper is

completed when the author is visiting IAS and he would like to thank IAS for its hospitality.

REFERENCES

- [BM] A. Bounemoura, J.-P. Marco, Improved exponential stability for near-integrable quasi-convex Hamiltonians, *Nonlinearity*, **24**, (2011), 97-112.
- [CCF] A. Córdoba, D. Córdoba, M. Fontelos, Formation of singularities for a transport equation with nonlocal velocity. *Annals of Mathematics*, **162**, (2005), 1377-1389.
- [FGKR] J. Féjoz, M. Guardia, V. Kaloshin, P. Raldan, Diffusion along mean motion resonance in the restricted planar three- body problem, <http://arxiv.org/abs/1109.2892>.
- [KNV] A. Kiselev, F. Nazarov, A. Volberg, Global well-posedness for the critical 2D dissipative quasi-geostrophic equation. *Inventiones Math.* **167**, (2007), 445–453.
- [L1] P. Lochak, Canonical perturbation theory via simultaneous approximation, *Russian Mathematical Surveys*, no. **6**, (1992), 57-133.
- [L2] P. Lochak, Simultaneous Diophantine approximation in classical perturbation theory: why and what for? *Progress in nonlinear science. Vol.1* RAS, Inst. Appl. Phys., Nizhnii (Novgorod), (2002), 116-138.
- [LN] P. Lochak, A.I. Neishtadt, Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian, *Chaos*, no. **4**, (1992), 495-499.
- [LNN] P. Lochak, A. I. Neishtadt, L. Niederman, Stability of nearly integrable convex Hamiltonian systems over exponentially long times. Kuksin, S. (ed.) et al., *Seminar on dynamical systems*. Basel: Birkhäuser. *Prog. Nonlinear Diff. Equ. Appl.* **12**, (1994), 15-34 .
- [N] L. Niederman, Stability over exponentially long times in the planetary problem, *Nonlinearity*, **9**, (1996), 1703-1751.
- [Ne] N. Nekhoroshev, An exponential estimate of the time of stability of nearly integrable Hamiltonian systems, *Russ. Math. Surv.* **32**, 1-65.
- [Po] J. Pöschel, Nekhoroshev estimates for quasi-convex Hamiltonian systems, *Mathematische Zeitschrift*, **213**, (1993), 187-216.
- [PT] A. Pronin, D. Treschev, Continuous averaging in multi-frequency slow-fast systems. *Regular and Chaotic Dynamics*, **5:2**, (2000), 157-170.
- [TZ] D. Treschev, Oleg Zubelevich. *Introduction to the perturbation theory of Hamiltonian systems*, Springer, (2009).
- [Tr1] D. V. Treschev, The continuous averaging method in the problem of separation of fast and slow motions. *Regular and Chaotic Dynamics*, **2:3/4**, (1997), 9-20.
- [Tr2] D. Treschev, Separatrix splitting for a pendulum with rapidly oscillating suspension point. *Russian J. Math. Phys.* **5:1**, (1997), 63-98.